

THE IMPACTS OF VOLUNTARY PRIVATE LANDS PROGRAMS ON STREAM FISH  
DIVERSITY IN THE KASKASKIA RIVER BASIN, ILLINOIS

BY

LEVI ALEXANDER DRAKE

THESIS

Submitted in partial fulfillment of the requirements  
for the degree of Master of Science in Natural Resources and Environmental Sciences  
in the Graduate College of the  
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Master's Committee:

Dr. Yong Cao, Co-advisor  
Dr. Leon Hinz, Co-advisor  
Associate Professor Cory Suski

## **Abstract**

Freshwaters support over 40% of fish species diversity, as well as one-third of all vertebrate species, yet remain one of the most threatened habitats globally. Anthropogenic disturbances have caused many negative impacts throughout history, and continue to do so today. After the dust bowl we began to inch our way toward smarter management of our watersheds. This eventually spurred the development of best management practices (BMPs) to combat non-point source pollution. Voluntary private lands programs such as the Conservation Reserve Program (CRP) look to offer monetary incentives to landowners willing to implement conservation practices on their lands. Biological goals, such as increased native bird or fish populations, are sometimes included in programs like CRP but little has been done to evaluate whether those goals are being achieved or not. Sampling can often be expensive for these endeavors, so alternative measures for obtaining this information are valuable. Species distribution modeling (SDM) has provided us with a chance to gain more information about communities without additional sampling effort. I look to balance sampling efforts with species distribution modeling to investigate the effects of CRP of stream fish species richness.

In this study, I use data from two Illinois fisheries datasets in combination with GIS environmental data to predict the presence or absence of 64 fish species across the Kaskaskia River basin using random forest classification. Of the 64 modeled species, 52 SDMs met my model performance requirements ( $TSS > 0.2$ ). These 52 SDMs were then stacked to obtain an index of species richness across the basin, and then the species richness values were compared with observed richness of modeled species, via regression, for accuracy. The regression deviated from the ideal 1:1 line, but Theil's Inequality Coefficient indicated a very strong matchup between observed and predicted richness ( $U = 0.012$ ). Based on this, I concluded that my SDMs

were able to provide a reasonable representation of species richness when the predictions of individual species models were stacked.

I then developed a novel standardization method using a house-neighborhood framework. “Neighborhoods”, all stream reaches within a given waterway distance from a site, were built around a group of fish sampling sites in the Kaskaskia River basin, Illinois. The species richness of the neighborhood was then used to standardize species richness at fish sampling sites. It is expected that a site in a neighborhood with high species richness would have more species than a site in a neighborhood with low species richness. Standardization based on the neighborhood species richness removes this species pool effect. Logit regression was then used to assess the effect of local habitat variables including CRP on species richness. Proportion of CRP lands within the local watershed for sampled sites ranged from 0% to 45.13%. Using the dredge function within the MuMIn package in R, all possible models were explored.  $R^2$  values were low across all models, ranging from  $R^2 = 0.0915$  to  $R^2 = 0.2367$ . The best models ( $\Delta AIC < 2$ ) took various combinations of in-stream habitat characteristics with large substrate consistently being ranked as one of the most important variables for species richness. The proportion of CRP lands in the local watershed was not taken as a predictor for any of the top models, while local habitat variables were found to be the most common factors influencing species richness. In conclusion, my study was unable to detect any major influence from CRP on stream fish species richness, and shows that local habitat factors are drivers of species richness when removing species pool effects from models. More rigorous targeting in the CRP implementation plans may help to increase the effect that CRP lands can have on fish species richness.

## **Acknowledgements**

I am so grateful for the support that I have received from so many amazing people throughout my studies. I would like to thank Dr. Peter Schroeder for being an incredible influence, and giving me my first introduction to the world of stream ecology, which has since become my passion. Next, Dr. Michael Parker helped me to develop and expand my knowledge of fisheries and aquatic ecology, and I may not be here today without his guidance. Most importantly, the experience and knowledge that was passed on to me from Dr. Yong Cao and Dr. Leon Hinz during my graduate career is priceless. Their guidance has transformed not only my education, but also my person. Thank you both so much.

Of course, this project would have never been possible without funding from the Illinois Department of Natural Resources. Also, I would like to thank Dr. Cory Suski for offering his thoughts and advice as my third committee member, and as a valued professor. Thank you to everyone in the INHS Stream Ecology Research Group, and I would like to give special thanks to Brian Metzke for the guidance and support he offered me during my few years here.

## Table of Contents

Chapter 1: Freshwaters and Conservation: A Review of Best Management Practices .....	1
Chapter 2: Fish Species Distribution Modeling in the Kaskaskia Basin using Random Forests Classification .....	15
Chapter 3: The Conservation Reserve Program Effects on Stream Fish Diversity in the Kaskaskia River Basin, IL .....	39
Chapter 4: Summary .....	63

## Chapter 1

### Freshwaters and Conservation: A Review of Best Management Practices

#### *Freshwater Ecosystems*

Freshwater represents only 2.5% of the water that covers more than 70% of the Earth's surface (McAllister et al. 1997). Moreover, only 0.01% of the Earth's water is found in lakes, rivers, and streams (Gleick 1996), leaving the rest of freshwater in underground aquifers or frozen in polar ice caps. This tiny fraction supports one third of all vertebrate species, including over 40% of fish species (Lundberg et al. 2000, Dudgeon et al. 2006). This represents the highest species richness relative to habitat extent in the world; greater than both marine and terrestrial systems (Silk and Ciruna 2013).

A diverse array of individual life cycles are found in freshwaters. Many examples exist of species whose life cycles depend on the seasonal variations and natural flood cycles that help to create greater habitat complexity and diversity year round. These species or populations have evolved to recognize seasonal cues, and one portion or another of their life cycles are affected or triggered by the seasonal changes to which freshwater ecosystems are prone. For example, diapause is common for insects and represents an adaptation to seasonal variation (Tauber and Tauber 1976, Denlinger 1986, Wolda 1988). Species migrations are another example of seasonally dependent life histories (Dingle 2014). When optimal environmental conditions occur at breeding sites, species migrations to these locations will tend to take place (Alerstam and Lindström 1990, Hodgson and Quinn 2002, Prop et al. 2003). Some salmonids exhibit these traits to enhance their reproductive success (Hodgson and Quinn 2002, Cooke et al. 2004,

Crossin et al. 2009). These traits contribute to the rich diversity found in freshwaters.

Maintaining freshwater species diversity will be imperative to future conservation.

Despite their importance to many organisms and their disproportionately high biodiversity, freshwater systems are threatened globally (Dudgeon et al. 2006, Silk and Ciruna 2013). They have experienced, and continue to experience, a large amount of anthropogenic disturbance which has resulted in reduced water quality, habitat loss, invasions by non-native species, and much more (Allan 2004, Dudgeon et al. 2006, Blann et al. 2009, Feld et al. 2011). Urbanization has been one form of disturbance that has taken its toll on freshwaters, as large expanses of impermeable surfaces which make up urban areas have amplified stormwater runoff into streams and rivers near cities (Paul and Meyer 2001, Feld et al. 2011). This runoff has been shown to greatly worsened water quality conditions in the areas downstream (Arnold Jr and Gibbons 1996, Booth and Jackson 1997). Dams have fragmented rivers due to water and power demand, as well as for flood control (Fuller et al. 2015, Grill et al. 2015, Zarfl et al. 2015). Agricultural lands have shown decreased water quality as well, due to anthropogenic disturbance. Some forestry practices have increased sediment loss to streams and decreased water quality (Broadmeadow and Nisbet 2004). Farming has impacted streams near farmlands through tile drainage and channelization, which have altered channel morphology and hydrology resulting in increased sedimentation and nutrient input to streams (Allan 2004, Blann et al. 2009). This is very apparent in streams in the Midwest, where these practices have been employed to remove water from poorly drained soils and efficiently export that water downstream (Randall et al. 1997, McIsaac and Hu 2004). Practices such as these have considerably altered stream flows, which has resulted in changes in both the timing and magnitude of natural, seasonal disturbances (Blann et al. 2009). With human-induced climate change looming, these alterations are only expected to

become more of an issue for both humans and stream dependent species(Ficke et al. 2007, Kundzewicz et al. 2008).

Ecologists have realized the importance of the roles that biodiversity and natural systems play, and particularly in freshwater systems for a long time (Dobson et al. 1997, Schwartz et al. 2000, Dudgeon et al. 2006). Biodiversity and ecosystem conservation efforts have been, and continue to be, carried out in order to preserve the goods and services that various ecosystems can provide. Efforts have been made in various ways, at scales ranging from local actions to global coordination (Brooks et al. 2006, Rodrigues 2006). In the United States, federal and regional programs implement best management practices (BMPs) on both private and public lands in an effort to mitigate some of our influence on aquatic systems (Ice 2004, Brady 2007, Knight and Boyer 2007). BMPs can take many forms, ranging from planting and tillage techniques to riparian protection or woody debris placement in streams. BMPs have become an important tool in our effort to take the steps necessary to restore natural ecological processes to systems in which these processes have been drastically altered (Ice 2004). They continue to be improved upon by managers and researchers alike to make their long-term effects as substantial as possible.

### *Best Management Practices*

A realized need for soil and water conservation grew following the dust bowl (Craddock and Hursh 1949). The creation of soil conservation districts and their ensuing management decisions laid the groundwork for modern BMPs (Ice 2004). Following a suite of federal acts and decisions throughout the mid-20<sup>th</sup> century (notably, the 1972 Federal Water Pollution Control Act, widely known as the Clean Water Act), BMPs became more refined in both scope and design. Non-point source pollution was defined as pollution not coming from a single known



pipe, ditch, or similar structure, and thus BMPs were focused to mitigate such pollution in accordance with United States water quality goals. The allure of BMPs is that they provide an opportunity to reduce human impacts on natural systems and encourage ecological function, while balancing the economic impact that could occur from taking land out of agricultural production (Ice 2004). This balance should theoretically help to ease any social tensions surrounding environmental restrictions and allow for society to reap benefits from the goods and services that a properly functioning ecosystem can provide. To achieve this balance, however, a mix of environmental, engineering, and economic expertise is required.

Water quality impacts are often the targets of BMPs. Practices are developed starting with knowledge of a water quality concern. Management choices are explored, and the practice that is deemed as best able to mitigate the concern is employed. Ideally, ongoing monitoring takes place to ensure that the goals of the management activity are met (Ice 2004). Monitoring efforts should attempt to take into account the lag time that exists between BMP implementation and the time when positive results may actually be seen (Meals et al. 2010). Eventually monitoring results can have the effect of encouraging improvement of the BMPs themselves over time and influencing the direction of environmental protections as their benefits or failures become more apparent. A proper monitoring strategy is essential to delivering effective impact mitigation.

In forestry, BMPs are utilized to protect streams from the negative water quality impacts that timber harvest can have (Ice 2004, Shepard 2006). These include increased temperature, decreased dissolved oxygen, and increased sediment transport. In addition, there are even BMPs designed to mitigate the effects of things such as logging road construction, road decommissioning, fire management, pesticide application, etc. Studies have shown that BMPs do

not remove all effects from silvicultural operations, but they do effectively reduce the impact to streams to only a small decrease in water quality from pre-harvest levels (Lynch and Corbett 1990, Williams et al. 1999). The exact reduction in water quality impact is variable depending on the region being studied and the practices being employed, but sites with BMPs have shown an 80-99% reduction in impact from sites without BMPs (Ice 2004). Discussion and application of forestry BMPs changes slightly depending on the region, but always maintains a core effort of careful planning, minimizing bare soil, revegetating quickly, and incorporating streamside management zones (Aust and Blinn 2004).

Agriculture is the leading source of impairment in streams and rivers in the United States (United States Environmental Protection Agency 2009). Much of the conservation effort in the agricultural realm is focused on private lands with voluntary implementation of BMPs by landowners. The voluntary nature of these programs offers a challenge, though, as many factors influence whether any particular landowner will adopt BMPs (Prokopy et al. 2008). Programs such as the Conservation Reserve Program (CRP) attempt to address this challenge by incentivizing conservation. Monetary compensation is offered to landowners for employing BMPs on their lands over a contracted period of time. Ideally, the collection of programs available to landowners for voluntary BMP implementation will ultimately result in the reduction of the effects seen in aquatic ecosystems from agricultural production.

However, voluntary BMP programs have shown very mixed results with regard to stream condition improvement and impact mitigation (Allan 2004, Liu et al. 2008, Tomer and Locke 2011, Rittenburg et al. 2015). Tomer and Locke (2011) suggested that a disconnect exists between practice-based studies and watershed scale studies, where it is not uncommon for practices to show improvement of water quality conditions but the watershed as a whole to still

show overall impairment. This is problematic, and may be contributed to a number of factors including the lag time for effects of conservation practices to be seen (Meals et al. 2010, Feld et al. 2011) or improperly targeted BMPs. Studies have indicated a need for better targeting procedures in BMP implementation so that areas that are more critical for conservation can receive more focus from conservation efforts (Tomer and Locke 2011, Holmes et al. 2016). This is an increasingly important realization, that the coordination and connectivity between the locations of BMPs are necessary to account for in both implementation and the assessment of their effects (Brueggen-boman et al. 2015, Brooks et al. 2015). While this may be difficult to achieve with voluntary programs, incorporating a more holistic view of watershed connectivity with BMP implementation may help attain much better results for impact mitigation.

The importance of connectivity is well known in the biological realm. Species pools and their limitations are an increasingly studied attribute of populations in aquatic ecosystems (Palmer et al. 1997, Sundermann et al. 2011, Stoll et al. 2013, 2014), and freshwater metacommunity dynamics are well explored (Erős et al. 2012, Heino 2013). Rivers and streams have a unique characteristic affecting their connectivity, which is their networked nature. At larger scales, it confines individuals to longitudinal movement along the stream network (Fagan 2002). This adds a level of complexity which increases the difficulty of trying to understand the nature of the relationships between populations and BMPs, habitat, or water quality across the stream network. Regardless, it is an important factor to include in any study looking to assess the effects of BMPs.

## *Objectives*

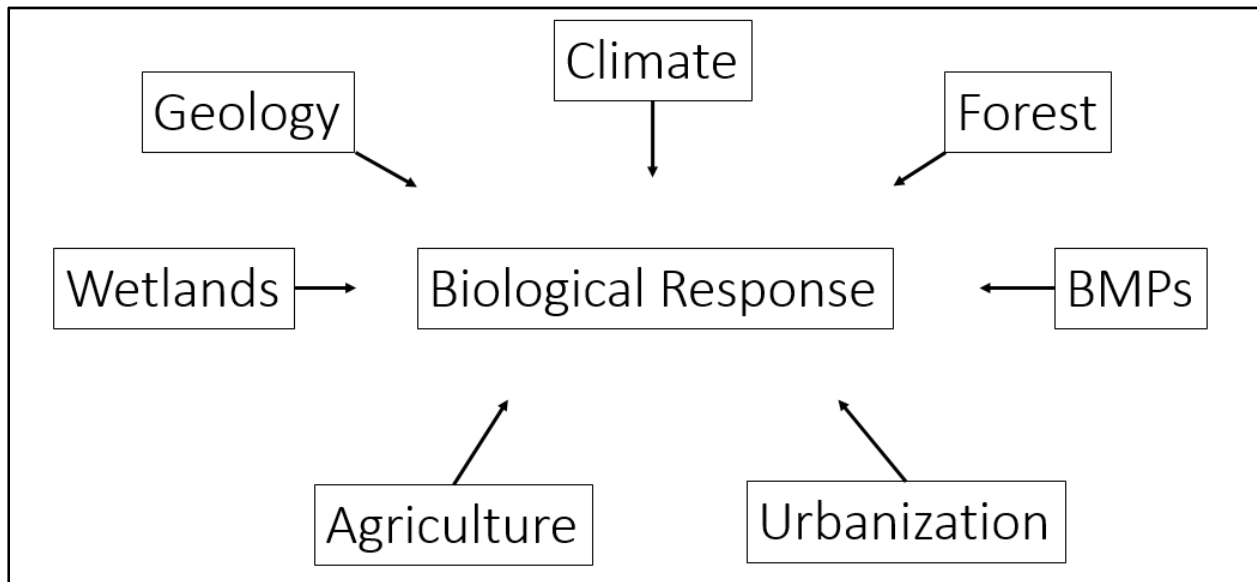
In this study I attempt to look at the effect of one BMP-based voluntary conservation program, CRP, on stream fish species richness. It is apparent that CRP has not been adequately studied in regard to its effectiveness, as applied to biological diversity in streams, in many places. In order to move forward with the biological goals for which CRP may be able to influence, and to not waste valuable time and money spent on public programs, then this relationship and its impact must be studied. My approach includes a combination of species modeling techniques and integrates connectivity through the incorporation of the regional species pool concept (Zobel et al. 1998, Dupré 2000). Numerous studies have described how land use affects streams, but a common issue for scientists and managers is the associated environmental noise that occurs when trying to look at relationships between land use variables and stream condition. When there is a whole watershed worth of geologic and land use variables interacting (Figure 1), we are presented with a big challenge to find an effect from a particular management activity. This has forced me to go to great lengths in this study in effort to reduce landscape influence in my models.

What I've done to minimize noise is look to the neighborhood species pool. I define the neighborhood species pool as the species inhabiting reaches at stream distances, or "neighborhoods", of 5km, 10km, 15km, and 20km away from a study location. For example, in Figure 2 the grey dot represents a study location and the neighborhoods are represented by the colored sections of stream. The richness in the neighborhood reflects the land use and natural environmental factors in the watershed and allows us to gain a better understanding of how reaches in the area might look given no local variability within the stream. The neighborhood richness is also the limiting richness at the study site, and incorporating it removes any effect that

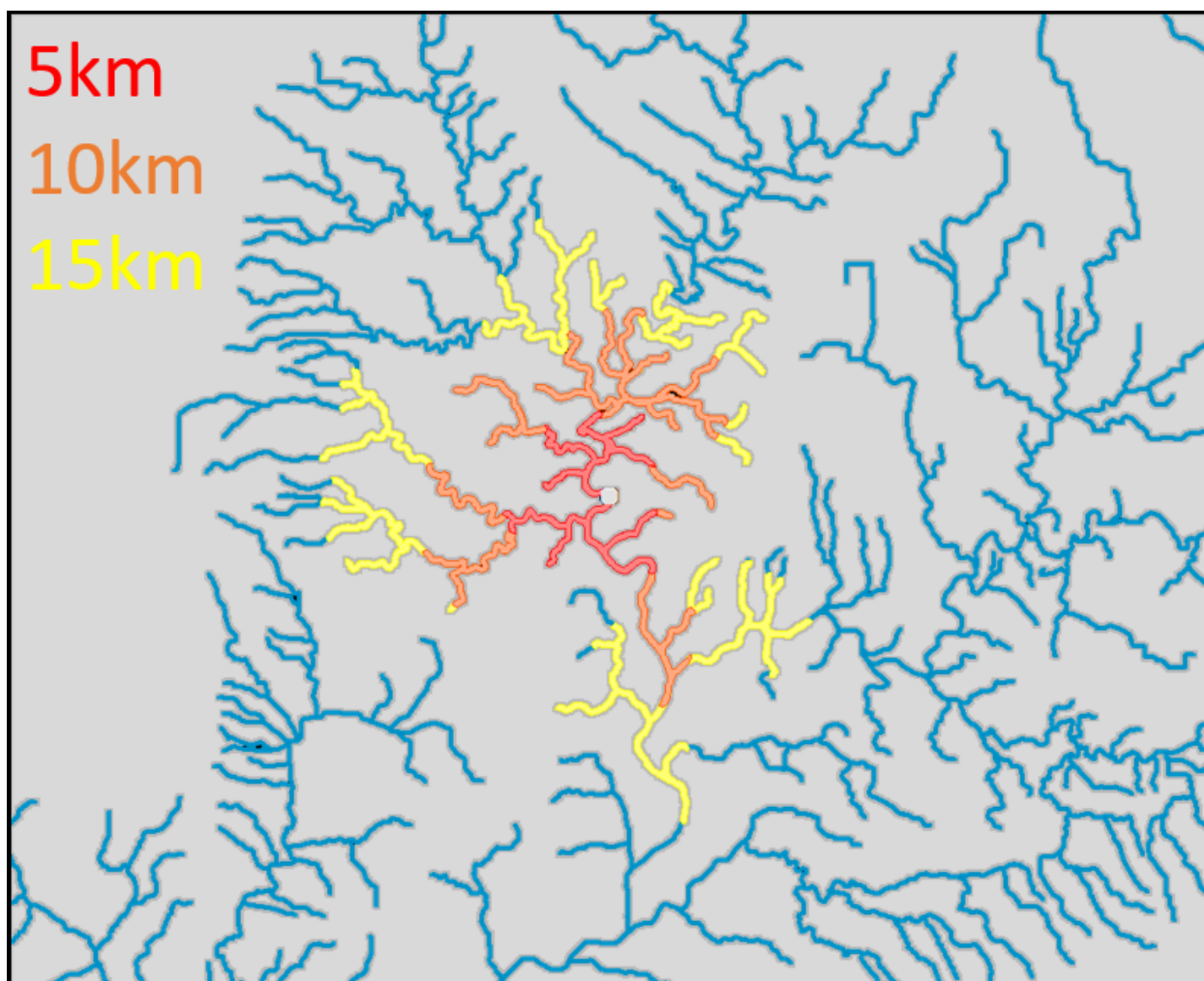
the species pool imposes on the species richness at the site. By standardizing our measured, site-specific species richness by the neighborhood richness, I can better isolate the effect of CRP management activities. I can then more closely examine the relationship between practices in the local watershed (the area that drains directly into a section of stream) and our measured richness.

Obtaining data for every section of stream 20km from a study location is time and cost prohibitive, however, so I have turned to species modeling. In Chapter 2, species modeling is used to obtain fish species presence/absence estimates, at a fine scale, over the entire Kaskaskia River basin in Illinois, which would be unfeasible to obtain by sampling alone. Modeling is performed for the basin using fisheries sampling data and environmental data which were compiled from multiple sources. Chapter 3 looks at the relationship between neighborhood standardized, fish species richness and local habitat variables, including the proportion of CRP lands in local watersheds. Chapter 4 will summarize findings and look to identify shortcomings and future research possibilities.

## Figures



**Figure 1.1** Conceptual representation of landscape variables that can affect a biological response in streams.



**Figure 1.2.** Hypothetical stream network showing neighborhoods (red, orange, and yellow) around a sampling site (grey dot at center).

## References

- Alerstam, T., and Å. Lindström. 1990. Optimal bird migration: the relative importance of time, energy, and safety. Pages 331–351 *Bird migration*. Springer.
- Allan, J. D. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35(1):257–284.
- Arnold Jr, C. L., and C. J. Gibbons. 1996. Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American planning Association* 62(2):243–258.
- Aust, W. M., and C. R. Blinn. 2004. Forestry best management practices for timber harvesting and site preparation in the eastern United States: An overview of water quality and productivity research during the past 20 years (1982–2002). *Water, Air and Soil Pollution: Focus* 4(1):5–36.
- Blann, K. L., J. L. Anderson, G. R. Sands, and B. Vondracek. 2009. Effects of Agricultural Drainage on Aquatic Ecosystems: A Review. *Critical Reviews in Environmental Science and Technology* 39(11):909–1001.
- Booth, D. B., and C. R. Jackson. 1997. Urbanization of aquatic systems: degradation thresholds, stormwater detection, and the limits of mitigation1. *Wiley Online Library*.
- Brady, S. J. 2007. Effects of Cropland Conservation Practices on Fish and Wildlife Habitat.
- Broadmeadow, S., and T. R. Nisbet. 2004. The effects of riparian forest management on the freshwater environment: a literature review of best management practice. *Hydrology and Earth System Sciences* 8(3):286–305.
- Brooks, E. S., S. M. Saia, J. Boll, L. Wetzel, Z. M. Easton, and T. S. Steenhuis. 2015. Assessing BMP Effectiveness and Guiding BMP Planning Using Process-Based Modeling. *JAWRA Journal of the American Water Resources Association* 51(2):343–358.
- Brooks, T. M., R. A. Mittermeier, G. A. B. da Fonseca, J. Gerlach, M. Hoffmann, J. F. Lamoreux, C. G. Mittermeier, J. D. Pilgrim, and A. S. L. Rodrigues. 2006. Global biodiversity conservation priorities. *science* 313(5783):58–61.
- Brueggen-boman, T. R., S. Choi, and J. L. Bouldin. 2015. Response of Water-Quality Indicators to the Implementation of Best-Management Practices in the Upper Strawberry River Watershed , Arkansas:697–713.
- Cooke, S. J., S. G. Hinch, A. P. Farrell, M. F. Lapointe, S. R. M. Jones, J. S. Macdonald, D. A. Patterson, M. C. Healey, and G. Van Der Kraak. 2004. Abnormal migration timing and high en route mortality of sockeye salmon in the Fraser River, British Columbia. *Fisheries* 29(2):22–33.
- Craddock, G., and C. Hursh. 1949. Watersheds and How To Care for Them. Pages 603–608 *Trees: 1949 Yearbook of Agriculture*. United States Department of Agriculture, Washington, D.C.



- Crossin, G. T., S. G. Hinch, S. J. Cooke, M. S. Cooperman, D. A. Patterson, D. W. Welch, K. C. Hanson, I. Olsson, K. K. English, and A. P. Farrell. 2009. Mechanisms influencing the timing and success of reproductive migration in a capital breeding semelparous fish species, the sockeye salmon. *Physiological and Biochemical Zoology* 82(6):635–652.
- Denlinger, D. L. 1986. Dormancy in tropical insects. *Annual review of entomology* 31(1):239–264.
- Dingle, H. 2014. *Migration: the biology of life on the move*. Oxford University Press, USA.
- Dobson, A. a. P., a. D. Bradshaw, and a. J. M. Baker. 1997. *Hopes for the future: restoration ecology and conservation biology*.
- Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. L&eacute;vy&ccedil;que, R. J. Naiman, A.-H. Prieur-Richard, D. Soto, M. L. J. Stiassny, and C. A. Sullivan. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* 81(02):163–182.
- Dupré, C. 2000. How to Determine a Regional Species Pool : A Study in Two Swedish Regions. *Oikos* 89:128–136.
- Erős, T., P. Sály, P. Takács, A. Specziár, and P. Bíró. 2012. Temporal variability in the spatial and environmental determinants of functional metacommunity organization – stream fish in a human-modified landscape. *Freshwater Biology* 57(9):1914–1928.
- Fagan, W. F. 2002. Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology* 83(12):3243–3249.
- Feld, C. K., S. Birk, D. C. Bradley, D. Hering, J. Kail, A. Marzin, A. Melcher, D. Nemitz, M. L. Pedersen, F. Pletterbauer, D. Pont, P. F. M. Verdonshot, and N. Friberg. 2011. *From Natural to Degraded Rivers and Back Again*. *Advances in Ecological Research*, 1st edition. Elsevier Ltd.
- Ficke, A. D., C. A. Myrick, and L. J. Hansen. 2007. Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries* 17(4):581–613.
- Fuller, M. R., M. W. Doyle, and D. L. Strayer. 2015. Causes and consequences of habitat fragmentation in river networks. *Annals of the New York Academy of Sciences* 1355(1):31–51.
- Gleick, P. H. 1996. Water resources. *Encyclopedia of climate and weather* 2:817–823.
- Grill, G., B. Lehner, A. E. Lumsdon, G. K. MacDonald, C. Zarfl, and C. R. Liermann. 2015. An index-based framework for assessing patterns and trends in river fragmentation and flow regulation by global dams at multiple scales. *Environmental Research Letters* 10(1):15001.
- Heino, J. 2013. The importance of metacommunity ecology for environmental assessment research in the freshwater realm. *Biological Reviews* 88(1):166–178.
- Hodgson, S., and T. P. Quinn. 2002. The timing of adult sockeye salmon migration into fresh water: adaptations by populations to prevailing thermal regimes. *Canadian Journal of Zoology* 80(3):542–555.

- Holmes, R., D. G. Armanini, and A. G. Yates. 2016. Effects of Best Management Practice on Ecological Condition : Does Location Matter ? Environmental Management.
- Ice, G. 2004. History of Innovative Best Management Practice Development and its Role in Addressing Water Quality Limited Waterbodies. *Journal of Environmental Engineering* 130(6):684–689.
- Knight, S., and K. Boyer. 2007. Effects of Conservation Practices on Aquatic Habitats and Fauna.
- Kundzewicz, Z. W., L. J. Mata, N. W. Arnell, P. Döll, B. Jimenez, K. Miller, T. Oki, Z. Şen, and I. Shiklomanov. 2008. The implications of projected climate change for freshwater resources and their management.
- Liu, X., X. Zhang, and M. Zhang. 2008. Major factors influencing the efficacy of vegetated buffers on sediment trapping: A review and analysis. *Journal of Environmental Quality* 37(5):1667–1674.
- Lundberg, J. G., M. Kottelat, G. R. Smith, M. L. J. Stiassny, and A. C. Gill. 2000. So Many Fishes, So Little Time: An Overview of Recent Ichthyological Discovery in Continental Waters. *Annals of the Missouri Botanical Garden* 87(1):26–62.
- Lynch, J. A., and E. S. Corbett. 1990. Evaluation of Best Management Practices for Controlling Nonpoint Pollution From Silvicultural Operations. *JAWRA Journal of the American Water Resources Association* 26(1):41–52.
- McAllister, D. E., B. Harvey, and A. L. Hamilton. 1997. Global freshwater biodiversity: striving for the integrity of freshwater ecosystems. Ocean Voice International.
- McIsaac, G. F., and X. Hu. 2004. Net N input and riverine N export from Illinois agricultural watersheds with and without extensive tile drainage. *Biogeochemistry* 70(2):253–273.
- Meals, D. W., S. A. Dressing, and T. E. Davenport. 2010. Lag Time in Water Quality Response to Best Management Practices: A Review. *Journal of Environmental Quality* 39.
- Palmer, M. a, R. F. Ambrose, and N. L. Poff. 1997. Ecological Theory and Community Restoration Ecology. *Restoration Ecology* 5(4):291–300.
- Paul, M. J., and J. L. Meyer. 2001. Streams in the Urban Landscape. *Annual Review of Ecology and Systematics* 32(1):333–365.
- Prokopy, L. S., K. Floress, D. Klotthor-Weinkauff, and a. Baumgart-Getz. 2008. Determinants of agricultural best management practice adoption: Evidence from the literature. *Journal of Soil and Water Conservation* 63(5):300–311.
- Prop, J., J. M. Black, and P. Shimmings. 2003. Travel schedules to the high arctic: barnacle geese trade-off the timing of migration with accumulation of fat deposits. *Oikos* 103(2):403–414.
- Randall, G. W., D. R. Huggins, M. P. Russelle, D. J. Fuchs, W. W. Nelson, and J. L. Anderson. 1997. Nitrate losses through subsurface tile drainage in conservation reserve program, alfalfa, and row crop systems. *Journal of Environmental Quality* 26(5):1240–1247.

- Rittenburg, R. A., A. L. Squires, J. Boll, E. S. Brooks, Z. M. Easton, and T. S. Steenhuis. 2015. Agricultural BMP Effectiveness and Dominant Hydrological Flow Paths: Concepts and a Review. *JAWRA Journal of the American Water Resources Association* 51(2):305–329.
- Rodrigues, A. S. L. 2006. Are global conservation efforts successful? *SCIENCE-NEW YORK THEN WASHINGTON* 313(5790):1051.
- Schwartz, M. W., C. a. Brigham, J. D. Hoeksema, K. G. Lyons, M. H. Mills, and P. J. van Mantgem. 2000. Linking biodiversity to ecosystem function: implications for conservation ecology. *Oecologia* 122(3):297–305.
- Shepard, J. P. 2006. Water quality protection in bioenergy production: the US system of forestry Best Management Practices. *Biomass and Bioenergy* 30(4):378–384.
- Silk, N., and K. Ciruna. 2013. A practitioner's guide to freshwater biodiversity conservation. Island Press.
- Stoll, S., J. Kail, A. W. Lorenz, A. Sundermann, and P. Haase. 2014. The importance of the regional species pool, ecological species traits and local habitat conditions for the colonization of restored river reaches by fish. *PloS one* 9(1):1–10.
- Stoll, S., A. Sundermann, A. W. Lorenz, J. Kail, and P. Haase. 2013. Small and impoverished regional species pools constrain colonisation of restored river reaches by fishes. *Freshwater Biology* 58(4):664–674.
- Sundermann, A., S. Stoll, and P. Haase. 2011. River restoration success depends on the species pool of the immediate surroundings. *Ecological Applications* 21(6):1962–1971.
- Tauber, M. J., and C. A. Tauber. 1976. Insect seasonality: diapause maintenance, termination, and postdiapause development. *Annual review of entomology* 21(1):81–107.
- Tomer, M. D., and M. A. Locke. 2011. The challenge of documenting water quality benefits of conservation practices: a review of USDA-ARSs conservation effects assessment project watershed studies. *Water Science and Technology* 64(1):300.
- United States Environmental Protection Agency. 2009. National Water Quality Inventory: Report to Congress. Washington, D.C.
- Williams, T. M., D. D. Hook, D. J. Lipscomb, X. Zeng, and J. W. Albiston. 1999. Effectiveness of best management practices to protect water quality in the South Carolina Piedmont. Pages 357–362 *Proceedings of the tenth biennial southern silvicultural research conference*. USDA Forest Service General Technical Report SRS-30T.
- Wolda, H. 1988. Insect seasonality: why? *Annual review of ecology and systematics*:1–18.
- Zarfl, C., A. E. Lumsdon, J. Berlekamp, L. Tydecks, and K. Tockner. 2015. A global boom in hydropower dam construction. *Aquatic Sciences* 77(1):161–170.
- Zobel, M., E. Van Der Maarel, C. Dupré, and V. Der Maarel. 1998. Species pool: the concept, its determination and significance for community restoration. *Applied Vegetation Science* 1(1):55–66.

## **Chapter 2**

### **Fish Species Distribution Modeling in the Kaskaskia Basin using Random Forests Classification**

#### **Abstract**

Data availability can prove to be a major limitation in ecology, as sampling can often be expensive. Advances in modeling techniques have provided us with useful tools to fill in data gaps and ask questions that we may not have been able to answer using sampling data alone. Species distribution models (SDMs) allow us to explore relationships between species and their environments and/or to predict the occurrence of a species at a location knowing only environmental data. Here, I use data from two Illinois fisheries datasets in combination with modeled environmental data to predict the presence or absence of 64 fish species across the Kaskaskia River basin using Random Forests classification. These predictions were stacked to establish neighborhood species pools later. Of the 64 modeled species, 52 SDMs met my model performance requirements ( $TSS > 0.2$ ). These 52 SDMs (average  $TSS = 0.4553$ ) were then combined additively to obtain species richness values across the basin, and then the species richness values were compared with observed richness, via regression, for accuracy. The regression deviated from the ideal 1:1 line, but Theil's Inequality Coefficient indicated a very strong matchup between observed and predicted richness ( $U=0.012$ ). Based on this, I concluded that my SDMs were able to provide an accurate representation of species richness when combined additively.

## **Introduction**

Modeling plays a major role in ecology. It allows us to predict future systematic changes based on past trends, or it can fill gaps in observed data sets. Both terrestrial and aquatic scientists take advantage of modeling techniques in their pursuits, where habitats, landscapes, and species distributions are all commonly modeled. These can be modeled independently or in tandem to answer ecological questions and further investigate the natural world.

Species distribution models (SDMs) are built using a combination of GIS derived environmental datasets and field-based species data (Elith and Leathwick 2009). Mobility of the species being modeled – whether sessile, like plants, or highly mobile, such as some birds or anadromous fish – is often a consideration and a factor influencing the modeling approach. Varying spatial scales are used, which are often determined by the scope and goals of the study. Some studies take a larger perspective looking at global or continental changes in species distributions (Rehfeldt et al. 2006, Lawler et al. 2009), while others take a more local or regional approach and look to explain particular patterns or solve local conservation issues (Hanski and Thomas 1994, Ferrier et al. 2002, Cao et al. 2015b, Schnier et al. 2016).

Often SDMs can be used for explanation and/or prediction (Mac Nally 2000). SDMs have been widely used to find a variable or suite of variables that explain species distributions (Lauenroth et al. 1993, Iriondo et al. 2003). Ecologists use them to uncover the processes that describe the patterns we see in ecological data. For prediction, SDMs have become more popular with the greater availability and sophistication of modeling methods in recent years (Elith and Leathwick 2009). Based on a set of training data, usually observed data from field ecology studies, predictive SDMs can either predict species occurrence at un-sampled locations within

the same geographical context as the training dataset (Labay et al. 2015), or they can make predictions for future changes in species distributions over time (Lawler et al. 2009).

Machine learning techniques are gaining traction in ecological modeling due to their flexibility and robust nature (Phillips et al. 2006, Hochachka et al. 2007). They are capable of handling large numbers of predictor variables, can identify the most important predictors, can predict with high accuracy, and do not require the user to make many unrealistic assumptions about the relationships between predictor variables and species. Random Forests (RF) classification (Breiman 2001) is one of these methods, which operates based on classification decision trees. A classification decision tree takes a large sample and breaks it apart into many subgroups using recursive binary partitioning. “Growing” this tree is done by splitting data based on predictors that best create two distinct “nodes.” This continues until a terminal node is reached and no further partitioning can be done, and a user defined number of observations occupy each node. Such a tree can create rules for assigning classifications to unsampled locations based on the set of predictors used to train the original tree.

RF takes the classification decision tree concept a step further by incorporating bootstrap sampling and ensemble methods. A randomly selected subset of the entire observed dataset is taken, consisting of two-thirds of the original data. The remaining observations are kept aside as an “out-of-bag” (OOB) sample, and a tree is grown from the subsample. A user-defined number of predictors (i.e. `mtry` in R package `randomForest`) are selected via bootstrap sampling, with replacement, from the total number of predictors for binary partitioning at each node in the tree. The predictor that best splits the data at a node is used. Each tree that is grown is applied to predict the observations in the OOB sample. An ensemble, or forest, is created by repeating this process many times, resulting in an ensemble of classification trees each made from a random

bootstrap sample. A class is predicted for an observation based on what the majority of trees concluded. Model performance can then be calculated based on the OOB predictions. A fully grown forest that is performing adequately can be used to predict classifications at unsampled locations with high accuracy. Cutler et al. (2007) provides an overview of RF application in ecology.

To further my study it is necessary to have reliable SDMs available for as many species as possible. SDMs are unavailable or incomplete for most fish species in the Kaskaskia basin, therefore SDMs need to be created. My objective is to build SDMs predicting presence/absence for individual stream fish species at each confluence-to-confluence reach in the Kaskaskia basin using RF classification. Individual model performance is evaluated, and models are also combined to find a predicted species richness at each reach. I then compare this predicted species richness to richness observations from another set of samples in the basin. Overall, this will provide individual species models that can be combined or used independently for further analyses.

## **Methods**

### *Study Area*

Illinois is highly impacted by human activities. Historically many parts of Illinois were covered with vast wet prairie, while woodland dominated the southern portion of the state (Sampson 1921). European settlement and agricultural advances transformed Illinois into an agriculturally dominated region – mostly corn and soybean agriculture – with small urban areas scattered across the landscape and very little original forest or native prairie remaining (Old

1969, Iverson 1988). Streams in this region have been highly altered as well. Rerouting and channelization have been utilized to remove water from agricultural fields more efficiently, but have resulted in the loss of instream and riparian habitat loss (Mattingly et al. 1993, Allan 2004, Blann et al. 2009). Tile drainage systems are also prevalent and add to the quick export of water from fields to streams, but at the expense of increased nitrogen export downstream (McIsaac and Hu 2004). A Mississippi River tributary in central and southern Illinois, the Kaskaskia River basin, is my area of focus in this study (Figure 2.1).

### *Fish Sampling and Spatial Data*

I compiled a dataset of electrofishing samples collected in the Kaskaskia basin from 1995 to 2015 to use in my modeling (Table 2.1). A total of 188 reaches were sampled during this period. These data were collected by Illinois Department of Natural Resources (IDNR) and Illinois Natural History Survey (INHS). Data from IDNR was from the Intensive Basin Survey, a joint statewide stream monitoring program between IDNR and Illinois Environmental Protection Agency (IEPA) started in 1981. Fish, aquatic macroinvertebrates, habitat, and water quality are all sampled under this program. Data was also collected from INHS Monitoring and Assessment of Aquatic Life in the Kaskaskia River for Evaluating IDNR Private Lands Programs (Hinz and Metzke 2015). Although the Intensive Basin Survey dataset goes back further than 1995, I used only the past 20 years of the dataset to reflect the current status of fish communities.

Fish samples under both programs were collected using standard IDNR protocols (IL-DNR 2010) by electric seine (Bayley et al. 1989) or backpack electrofishing, as well as boat electrofishing for some samples under the IDNR Intensive Basin Survey. Sampling reaches were



minimum 100 meters, 20x the stream width, and no more than 300 meters long. Reaches started or ended at least 10 meters above or below a road crossing to avoid any effects the bridge may have. Electric seine was used whenever width and water levels were great enough to merit usage, and the backpack shocker was used in only the smallest of sites.

The Great Lakes Regional Aquatic GAP Analysis from the United States Geological Survey (<http://gapanalysis.usgs.gov/aquatic-gap/>) modeled surficial environmental data for Great Lakes states which provided me with landscape environmental data for the Kaskaskia basin and a spatial framework to use in species modeling (Brenden et al. 2008, Steen et al. 2008, McKenna Jr and Johnson 2011, Cao et al. 2015a). This GIS dataset used by Cao et al. (2015a) contains 68 environmental variables attributed to each confluence-to-confluence section of stream at the 1:100,000 scale. Four spatial extents (total watershed, local watershed, 30m riparian zone of the total watershed, 30m riparian zone of the local watershed) were used to define environmental variables describing climate, geology, and landscape. These environmental variables were combined with the fisheries sampling dataset to create my full dataset for modeling presence and absence of individual species. The modeling dataset had 142 fish sampling locations from across the basin, and the remaining fish sampling data – 46 samples – were set aside as another dataset for model evaluation. I defined my study units at the same resolution as the environmental dataset, confluence-to-confluence sections of stream at the 1:100,000 scale.

### *Species Modeling*

To be included in my modeling, a species had to have been recorded in at least 10 different sampling locations in my dataset. This gives the RF classification models a set of

environmental factors from which the presence of a species can be predicted at other locations. A total of 64 species met this criterion. These 64 species were modeled for presence and absence using RF classification (Breiman 2001). Modeling was performed within R (R Core Team 2013), using the package “randomForest” (Liaw and Wiener 2002). Sensitivity, specificity, and true skill score (TSS)(Allouche et al. 2006) were calculated for every model. TSS ranges from -1 to 1 and is a measure of how well the model is able to correctly predict presence (sensitivity) or absence (specificity) of a species. Equations (1) through (3) demonstrate the calculation of TSS:

$$Sensitivity = \frac{d}{c+d} \quad (1)$$

$$Specificity = \frac{a}{a+b} \quad (2)$$

$$TSS = (Sensitivity + Specificity) - 1 \quad (3)$$

Definitions of variables in equations (1) and (2) can be found in the contingency table (Table 1). A value of TSS = -1 means that the model was unable to predict presence correctly anywhere and predicted presence at every site where the species was absent, while a value of TSS = 1 means that the model was able to predict both presence and absence with perfect accuracy. A TSS = 0 would mean that the model is performing as well as random assignment of presence and absence of the species. The literature does not suggest an accepted value for TSS that shows that an RF classification model is performing adequately. I chose TSS = 0.2 as my benchmark for useful model performance. The selection of TSS = 0.2 was made in an attempt to have models perform accurately more often than not, while simultaneously maximizing the number of species models able to be used in future analyses.

For each species, 5000 trees were grown and averaged using RF classification. This was done five separate times for every mtry – the number of variables used to split nodes in the

classification tree – between 2 and 10, leaving us with 45 different models for each species. The model with the highest TSS was selected as the best model, with the model using the lowest number of variables to split nodes selected in the case of ties. SDMs that were below the model performance cutoff were excluded from further analysis. The models that made the cut were combined additively to find predicted species richness at the 46 reaches that were not included in my original modeling dataset. Observed species richness ( $SR_{obs}$ ) that had been modified to include only species that had met my modeling criteria were compared to predicted species richnesses ( $SR_{pred}$ ). Linear regression was conducted in the OP arrangement recommended in Piñeiro et al. (2008), and  $R^2$  and Theil's Inequality Coefficient (U) were calculated. Theil's Inequality Coefficient is similar to a correlation coefficient, but instead of comparing distances from the regression line it incorporates the slope of the line and allows for a comparison to the ideal agreement that would be seen at the 1:1 line (Smith and Rose 1995). It ranges from 0 to 1, with 0 indicating an exact matchup. A ratio of  $SR_{obs}/SR_{pred}$  was also examined to gain another look at how species richness predictions compared to observed richness.

## Results

Overall, model performance was good. For RF classification modeling, the average TSS of all 64 SDMs was  $TSS = 0.3842$ . However, 12 species did not meet our minimum model performance of  $TSS = 0.2$  (Table 2). The average TSS of the remaining 52 SDMs above our model performance cutoff was  $TSS = 0.4553$ . The 12 species that were excluded ended up being species that were either very common, such as green sunfish (*Lepomis cyanellus*), or very rare, such as bowfin (*Amia calva*).

The information provided by the ratio for comparison ( $SR_{obs}/SR_{pred}$ ) was decent as well. The ratio ranged from a minimum of 0.200 to a maximum of 1.778, although one major outlier (3.500) from the distribution existed markedly beyond the outer fence of 2.363 (Figure 2). After removing the outlier, the mean of the distribution was 1.011 with a standard deviation of 0.346. If the major outlier is not removed, the mean of the distribution increases to 1.064 with a standard deviation of 0.496. The regression of  $SR_{obs}$  on  $SR_{pred}$  (Figure 3) produced a line  $SR_{obs} = 0.7181(SR_{pred}) + 3.023$  with  $R^2 = 0.4671$  and  $U = 0.011$ . Including the major outlier in the regression changes very little, producing a line  $SR_{obs} = 0.6485(SR_{pred}) + 4.006$  with  $R^2 = 0.4056$  and  $U = 0.012$  (Figure 4). Based on this, richness was slightly underestimated at sites with lower observed richness while the opposite was true for sites with higher observed richness.

## Discussion

The modeling done in this study was necessary to obtain a sufficient amount of data so that neighborhood species richness could be calculated around sampling sites in the next chapter. Without species modeling, obtaining this data would take far too much time, resulting in a cost-prohibitive study. It is important that these SDMs be as accurate as possible so that my calculations of the neighborhood species richness were reliable in future assessments.

As stated previously in the methods, current literature does not suggest a minimum TSS that indicates a model is performing adequately. Therefore it is up to authors to decide an acceptable model performance cutoff for the study at hand. With my TSS selection, the species that did not meet the cutoff were most often species that are rare in the Kaskaskia River basin or extremely abundant. This would make sense given the nature of the TSS calculation. For

example, species such as green sunfish are so abundant in the Kaskaskia basin that the conditions in which they exist can be incredibly diverse. This would likely lead the model to predict its presence at almost every reach in the basin, leading sensitivity for this model to be near one and specificity to be near negative one. The opposite would be true for extremely rare species, but both instances lead to a TSS at or near zero.

The division imposed by the cutoff is important, as I intended to combine these SDMs in aggregate to look at species richness using the concept of “predict first, assemble later” (Ferrier and Guisan 2006). Conceptually, the rarest species and the most abundant species are not typically the species driving changes in richness. These species are either contributing +1 toward richness constantly across most locations, or are consistently not contributing at all. Removing these species from the assessment can make sense, so as to focus on the species that more commonly drive the changes seen in richness across the basin.

The method I used to estimate species richness from individual SDMs is sometimes called stacked species distribution modeling (S-SDM) or stacking (Ferrier and Guisan 2006, Mateo et al. 2012). Current literature on this technique indicates that stacked models will tend to overestimate species richness (Guisan and Rahbek 2011, Dubuis et al. 2011, Mateo et al. 2012, Calabrese et al. 2013). For example, Labay et al. (2015) observed this trend at nearly all sites, but the difference was most extreme in species-rich areas. Calabrese et al. (2013) attributed the consistent overestimation of species richness in S-SDMs to incorrectly stacked SDMs based on threshold methods. My study stands contrary to this, in that my TSS threshold-based SDMs provided richness predictions that resembled those of the richness observed at those locations. Qualitatively, the fairly normal distribution around 1.000 shown in Figure 2 indicates the ability of the individual SDMs to be combined additively and still provide fairly accurate estimates of

richness. A distribution around 1.000 means that the models, when combined, are maintaining a relationship close to the observed richness. More importantly, Theil's Inequality Coefficient was  $U = 0.011$ , indicating a strong agreement between observations and predictions (a value of 0 indicates a theoretically perfect relationship on the scale from 0 to 1). This U-value is of interest for goodness of fit in this assessment because of its ability to include slope and compare to the ideal 1:1 relationship, rather than strictly observing whether the fit was good at any slope (Smith and Rose 1995, Piñeiro et al. 2008, Labay et al. 2015).

The conversation about overestimation of richness in S-SDMs originated due to discrepancies between their predictions and the predictions derived from macroecological models (MEM) of richness (Guisan and Rahbek 2011, Calabrese et al. 2013). MEM methods directly model species richness and have generally provided reliable estimates, however they do exhibit tendencies to overestimate at low richness locations and underestimate at high richness locations. Calabrese et al. (2013) noted that properly stacked SDMs will exhibit similar tendencies to MEMs. My models again exhibit different characteristics. Estimations of richness did deviate from the theoretically perfect 1:1 relationship, but underestimation was seen at low richness locations while overestimation was seen at higher richness locations.

Modeling individual species had two main applications in my study. Due to time limitations, I was not able to visit sites repeatedly for multiple sampling events. If I had, I would expect that variations in observed richness would likely occur. This would give me the ability to find the average species richness that would typically exist within a reach. However, it can be argued conceptually that richness derived from S-SDMs can estimate this number without the additional sampling effort at a single site. This describes my first reason for using S-SDMs. The richness defined by these models is developed based on conditions expected for individual

species to be present, and from numerous events from across the basin. This decreases the likelihood that a stray or wandering individual of a species not normally found in a particular location may be included in the richness estimate, while increasing the possibility of including a more difficult to detect species. Without these models, repeated sampling efforts would be needed to obtain this value, which results in more time and money on the part of the research group.

A second application for using individual SDMs is the amount of information they can provide. Individual SDMs can be combined in many ways to build a number of different community metrics (Ferrier and Guisan 2006). This increased information is valuable, and allows us the flexibility to not only use these models as predictors of richness but also the possibility for future assessments exploring distributions of particular focal species or community types. Also, Franklin (1995) argues that individual SDMs provide us with fewer ecological uncertainties than would modeling communities directly. Directly modeling species richness would be an example of a community based or macroecological model. While this could provide us with acceptable depictions of diversity, it would provide much less information due to certain inherent ecological assumptions they make. For example, communities do not necessarily move as entities in the face of disturbance due to different tolerance levels among species to disturbance type or magnitude (Huntley 1991, Williams and Jackson 2007). This lends itself to the idea that S-SDMs may offer us more reliable representations of the overall assemblage once they are combined. As they are built species-by-species, individual SDMs may be more responsive to the conditions that are driving presence and absence, and ultimately richness. It is important to note, though, that the individualistic nature of S-SDMs does not incorporate ecological constraints such as carrying capacity which can theoretically lead to overestimation of

richness (Guisan and Rahbek 2011). This reinforces the need for robust methods for providing accurate SDMs.

The use of Random Forests classification in this study is important. Robust machine learning techniques such as Random Forests are being used more and more in fisheries modeling (e.g. He et al. 2010, Knudby et al. 2010, Parker et al. 2015, Cao et al. 2015a), and my study is yet another example of its usefulness and accuracy in making predictions. Ultimately, in this study I have created SDMs using a proven, robust machine learning technique and tested the dependability of the SDMs. I showed that I could stack these SDMs and achieve a reasonable calculation of species richness from them, even using threshold-based techniques. My intent is to use these individual species models to define neighborhood species pools at multiple sampling locations in the Kaskaskia River basin. The strong matchup between my predicted species richness and the observed species richness, represented by Theil's Inequality Coefficient, makes me confident that this modeling effort has provided me with the necessary species distribution information to build the neighborhoods I require to assess the effect of local habitat factors and CRP lands on stream fish species richness.



## Tables and Figures

**Table 2.1.** Summary of data compiled from 1995-2015 Illinois fisheries datasets.

<i>Measure</i>	<i>Min</i>	<i>Mean</i>	<i>Max</i>
Species Richness	0	17.43	40
Stream Order	1	3.44	7
Stream Link	1	106.25	1928
Total Watershed Size (km <sup>2</sup> )	2.37	704.98	11992.51
Local Watershed Size (km <sup>2</sup> )	0.01	5.53	46.01
Total Watershed Forest (%)	0	8	31
Local Watershed Forest (%)	0	22	1
Total Watershed Agriculture (%)	32	71	99
Local Watershed Agriculture (%)	0	51	97
Total Watershed Slope (%)	0.01	0.58	2.41

**Table 2.2.** Model performance contingency table for sensitivity, specificity, and TSS calculation equations.

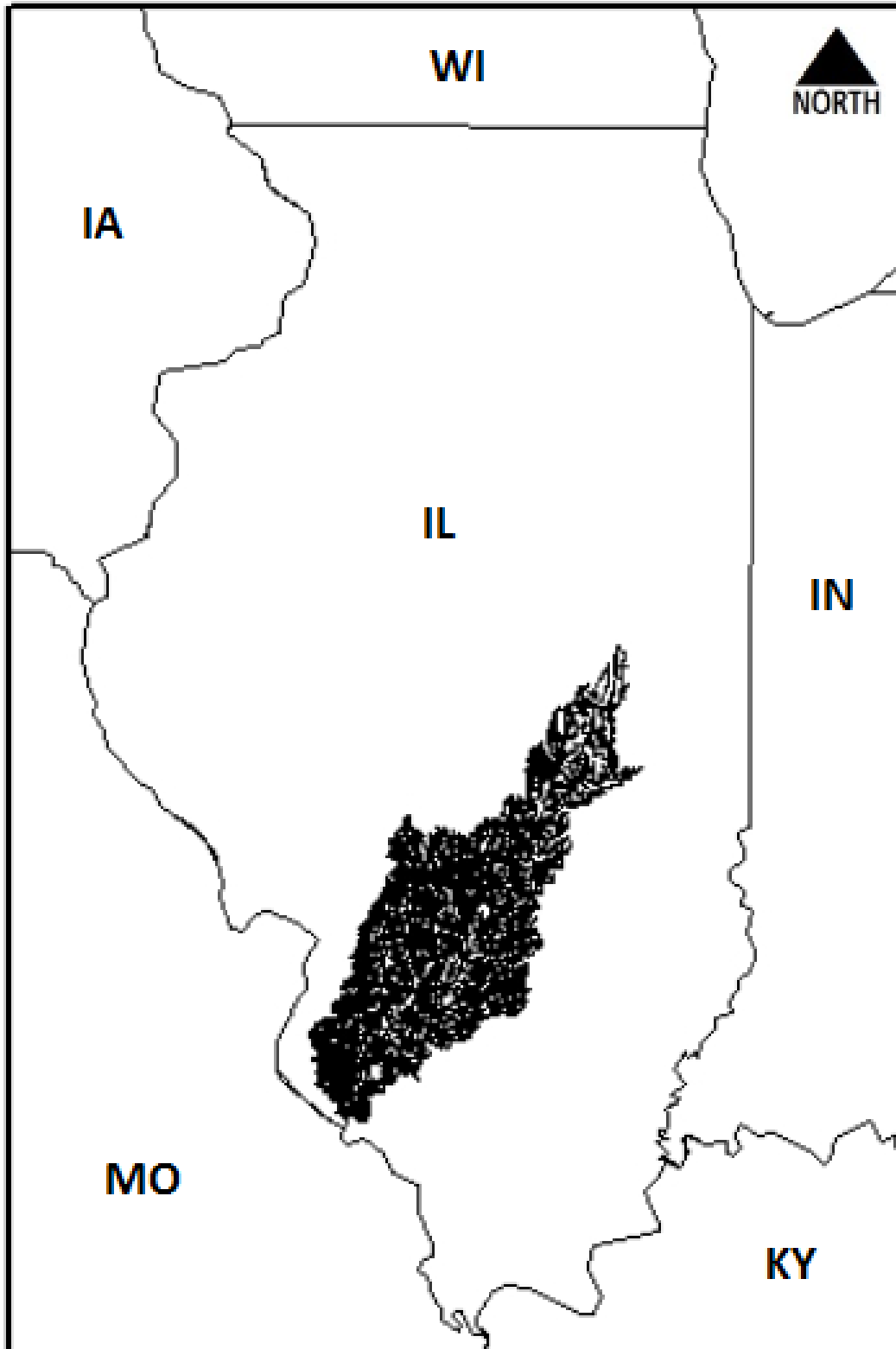
		<i>Predicted</i>	
<i>Observed</i>	<b>Absence</b>	<b>Absence</b>	<b>Presence</b>
	<b>Presence</b>	a: true absence c: false negative	b: false positive d: true presence

**Table 2.3.** Species modeled for presence/absence and their associated model performance metrics. Sensitivity, specificity, and true skill score (TSS) are as defined in the text; mtry represents the number of variables selected at each node of a random forests classification tree. \*Species denoted with an asterisk were excluded from the rest of our study due to poor model performance.

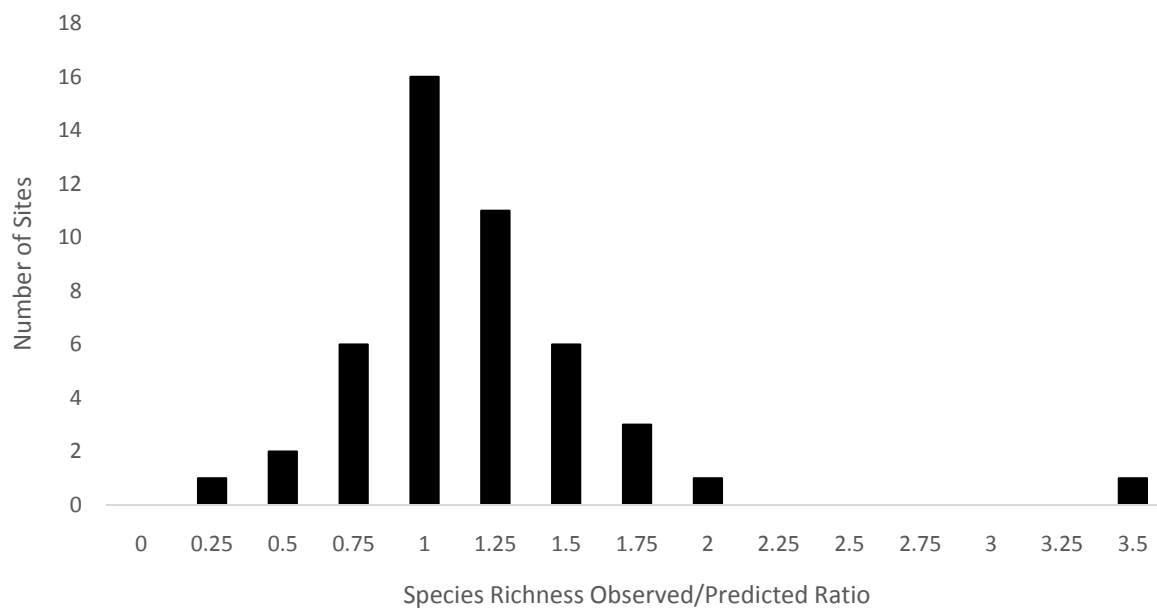
<i>Common Name</i>	<i>mtry</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>TSS</i>
Bigmouth Buffalo ( <i>Ictiobuscyprinellus</i> )	2	0.3214	0.9912	0.3127
Bluegill x Green Sunfish ( <i>Lepomis macrochirus</i> x <i>L. cyanellus</i> )*	2	0.0000	1.0000	0.0000
Black Buffalo ( <i>Ictiobus niger</i> )	8	0.4118	0.9680	0.3798
Black Bullhead ( <i>Ameiurus melas</i> )*	9	0.1579	0.9756	0.1335
Black Crappie ( <i>Pomoxis nigromaculatus</i> )	6	0.3600	0.9573	0.3173
Blackside Darter ( <i>Percina maculate</i> )	4	0.4865	0.9333	0.4198
Bluegill ( <i>Lepomis macrochirus</i> )	5	0.9364	0.3125	0.2489
Bluntnose Minnow ( <i>Pimephales notatus</i> )*	3	1.0000	0.0370	0.0370
Blackstripe Topminnow ( <i>Fundulus notatus</i> )	9	0.9358	0.2727	0.2085
Bigmouth Shiner ( <i>Notropis dorsalis</i> )	5	0.5789	0.9712	0.5501
Bowfin ( <i>Amiocalva</i> )*	2	0.0000	1.0000	0.0000
Brook Silverside ( <i>Labidesthes sicculus</i> )	3	0.6087	0.9916	0.6003
Bullhead Minnow ( <i>Pimephales vigilax</i> )	4	0.6829	0.9604	0.6433
Common Carp ( <i>Cyprinus carpio</i> )	3	0.7361	0.8000	0.5361
Channel Catfish ( <i>Ictalurus punctatus</i> )	3	0.7000	0.9457	0.6457
Creek Chubsucker ( <i>Erimyzon oblongus</i> )	5	0.5102	0.8602	0.3704
Central Stoneroller ( <i>Camptostoma anomalum</i> )	5	0.7662	0.7385	0.5047
Creek Chub ( <i>Semotilus atromaculatus</i> )	4	0.8947	0.5106	0.4054
Flathead Catfish ( <i>Pylodictis olivaris</i> )	7	0.5385	0.9569	0.4954
Fathead Minnow ( <i>Pimephales promelas</i> )*	2	0.0000	1.0000	0.0000
Freshwater Drum ( <i>Aplodinotus grunniens</i> )	3	0.6122	0.9247	0.5370
Freckled Madtom ( <i>Noturus nocturnus</i> )	7	0.4688	0.9455	0.4142
Golden Redhorse ( <i>Moxostoma erythrurum</i> )	5	0.4375	0.9545	0.3920
Golden Shiner ( <i>Notemigonus crysoleucas</i> )	7	0.2703	0.9429	0.2131
Grass Pickerel ( <i>Esox americanus</i> )	5	0.4848	0.9358	0.4206
Green Sunfish ( <i>Lepomis cyanellus</i> )*	3	1.0000	0.1429	0.1429
Gizzard Shad ( <i>Dorosomacepedianum</i> )	2	0.7121	0.8684	0.5805
Highfin Carpsucker ( <i>Carpiodes velifer</i> )	3	0.4615	0.9845	0.4460
Hornyhead Chub ( <i>Nocomis biguttatus</i> )	5	0.6000	0.9426	0.5426
Johnny Darter ( <i>Etheostoma nigrum</i> )	3	0.7945	0.7536	0.5481
Longear Sunfish x Green Sunfish ( <i>L. megalotis</i> x <i>L. macrochirus</i> )*	2	0.0000	1.0000	0.0000
Largemouth Bass ( <i>Micropterus salmoides</i> )	8	0.9434	0.3611	0.3045
Longnose Gar ( <i>Lepisosteus osseus</i> )	5	0.6000	0.9843	0.5843
Logperch ( <i>Percinacaprodes</i> )	5	0.2333	0.9732	0.2065
Longear Sunfish ( <i>Lepomis megalotis</i> )	5	0.9362	0.6875	0.6237
Western Mosquitofish ( <i>Gambusia affinis</i> )*	3	0.3962	0.7865	0.1827
Mud Darter ( <i>Etheostoma asprigene</i> )*	9	0.1053	0.9919	0.0971

**Table 2.3.** Continued.

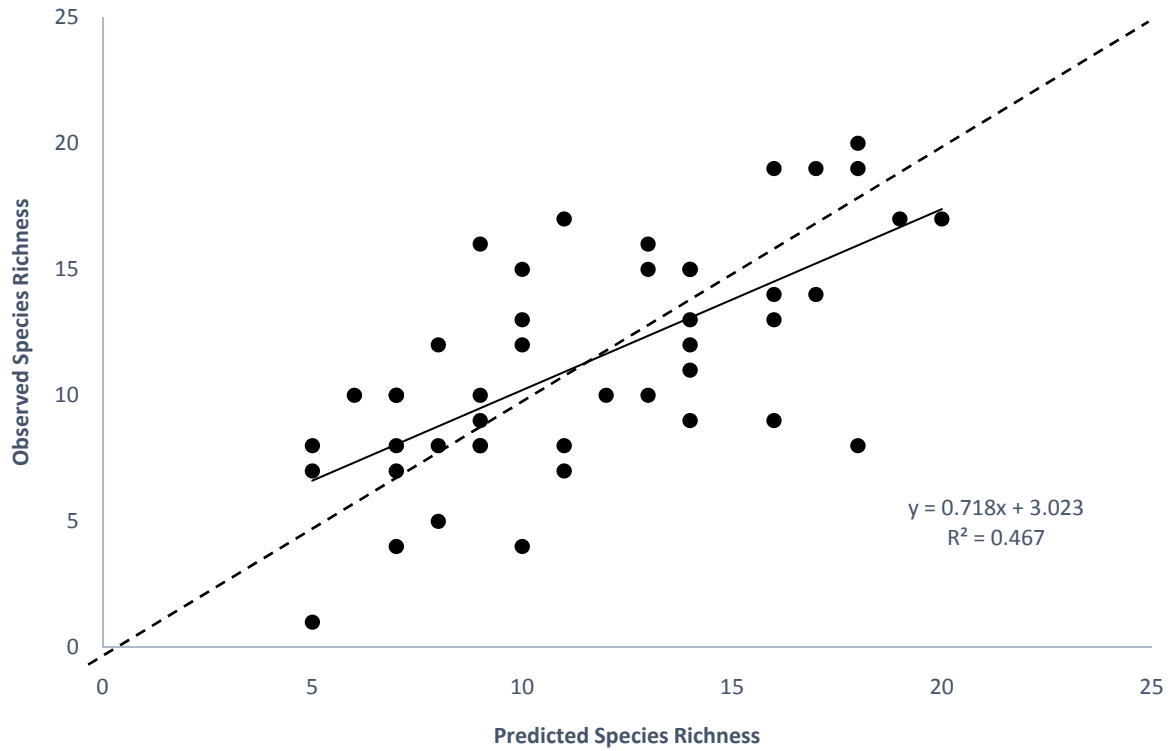
<i>Common Name</i>	<i>mtry</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>TSS</i>
Orangethroat Darter ( <i>Etheostomaspectabile</i> )	4	0.5172	0.9735	0.4907
Orange Spotted Sunfish ( <i>Lepomishumilis</i> )*	7	0.1613	0.9640	0.1253
Pirate Perch ( <i>Aphredoderussayanus</i> )	3	0.6552	0.8095	0.4647
Redfin Shiner ( <i>Lythrurusumbratilis</i> )	4	0.7260	0.6957	0.4217
Red Shiner ( <i>Cyprinellalutrensis</i> )	7	0.9252	0.3429	0.2681
Redear Sunfish ( <i>Lepomismicrolophus</i> )*	2	0.0000	1.0000	0.0000
River Carpsucker ( <i>Carpiondescarpio</i> )	2	0.6061	0.9725	0.5785
Smallmouth Buffalo ( <i>Ictiobusbubalus</i> )	4	0.6667	0.9826	0.6493
Sauger ( <i>Stizostedioncanadense</i> )	5	0.6154	0.9767	0.5921
Sand Shiner ( <i>Notropisludibundus</i> )	6	0.8101	0.7937	0.6038
Silver Carp ( <i>Hypophthalmichthysmolitrix</i> )	3	0.4286	0.9926	0.4212
Spotted Sucker ( <i>Minytremamelanops</i> )	5	0.2500	0.9649	0.2149
Slenderhead Darter ( <i>Percinaphoxocephala</i> )	4	0.4167	0.9576	0.3743
Shortnose Gar ( <i>Lepisosteusplatostomus</i> )	4	0.5556	0.9919	0.5475
Shorthead Redhorse ( <i>Moxostomamacrolepidotum</i> )	4	0.7105	0.9615	0.6721
Silverjaw Minnow ( <i>Notropisbuccatus</i> )	4	0.6780	0.9036	0.5816
Slough Darter ( <i>Etheostomagracile</i> )*	7	0.2174	0.9748	0.1922
Striped Shiner ( <i>Luxiluschrysocephalus</i> )	2	0.6786	0.9561	0.6347
Suckermouth Minnow ( <i>Phenacobiusmirabilis</i> )	3	0.7705	0.8642	0.6347
Tadpole Madtom ( <i>Noturusgyrinus</i> )	5	0.5424	0.7952	0.3376
Quillback ( <i>Carpiondescyprinus</i> )	2	0.5000	0.9327	0.4327
Warmouth ( <i>Lepomisgulosus</i> )	6	0.2778	0.9758	0.2536
White Bass ( <i>Moronechrysops</i> )	7	0.5455	0.9667	0.5121
White Crappie ( <i>Pomoxisannularis</i> )	3	0.4667	0.9175	0.3842
White Sucker ( <i>Catostomuscommersoni</i> )	5	0.7639	0.7571	0.5210
Yellow Bullhead ( <i>Ameiurusnatalis</i> )	5	0.9412	0.4000	0.3412
Yellow Bass ( <i>Moronemississippiensis</i> )	7	0.3333	0.9606	0.2940



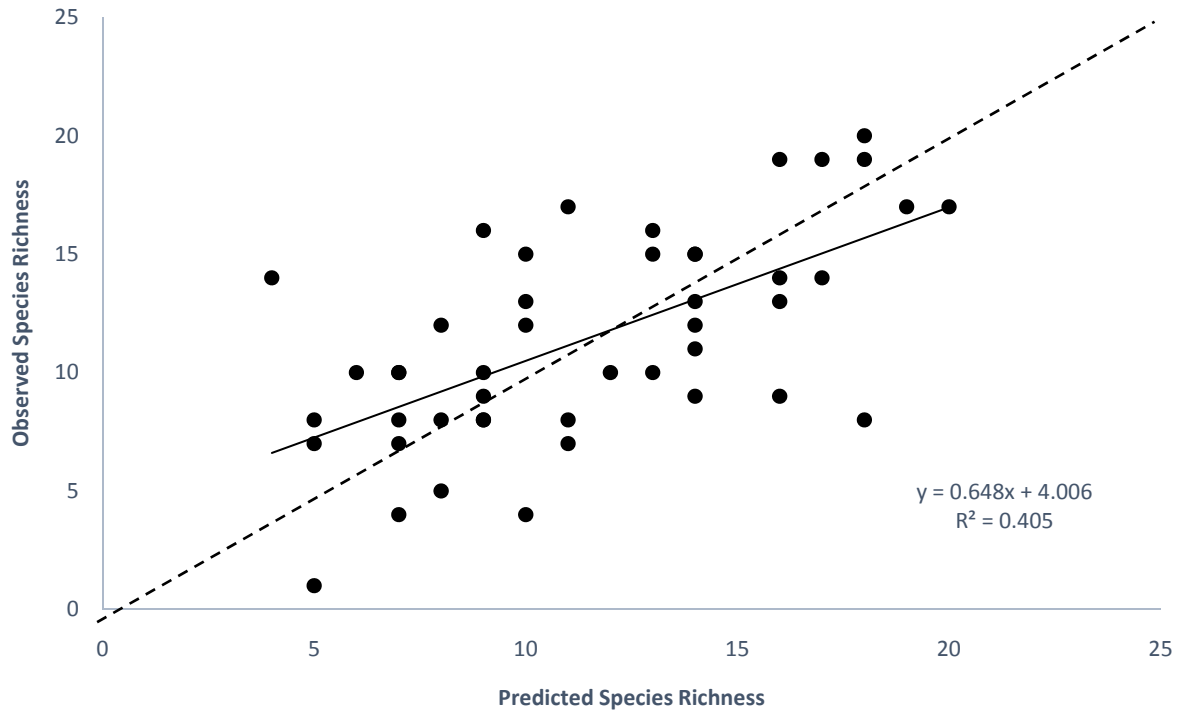
**Figure 2.1.** Geographical location of the Kaskaskia River Basin within Illinois, US.



**Figure 2.2.** Frequency histogram for the observed/predicted ratio for evaluating model performance. A value of 1 indicates a perfect match.



**Figure 2.3.** Observed vs Predicted correlation for model performance evaluation, excluding the major outlier. Axes are arranged according to suggestions by Piñeiro et al. (2008). Theil's Inequality Coefficient for this regression is  $U = 0.011$ . The dotted line represents the ideal 1:1 matchup line.



**Figure 2.4.** Observed vs Predicted correlation for model performance evaluation, including the major outlier. Axes are arranged according to suggestions by Piñeiro et al. (2008). Theil's Inequality Coefficient for this regression is  $U = 0.012$ . The dotted line represents the ideal 1:1 matchup line.

## References

- Allan, J. D. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35(1):257–284.
- Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43(6):1223–1232.
- Bayley, P. B., R. W. Larimore, and D. C. Dowling. 1989. Electric Seine as a Fish-Sampling Gear in Streams. *Transactions of the American Fisheries Society* 118(4):447–453.
- Blann, K. L., J. L. Anderson, G. R. Sands, and B. Vondracek. 2009. Effects of Agricultural Drainage on Aquatic Ecosystems: A Review. *Critical Reviews in Environmental Science and Technology* 39(11):909–1001.
- Breiman, L. 2001. Random Forests. *Journal of Machine Learning* 45(1):5–32.
- Brenden, T. O., L. Wang, P. W. Seelbach, R. D. Clark, M. J. Wiley, and B. L. Sparks-Jackson. 2008. A spatially constrained clustering program for river valley segment delineation from GIS digital river networks. *Environmental Modelling & Software* 23(5):638–649.
- Calabrese, J. M., G. Certain, C. Kraan, and C. F. Dormann. 2013. Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography* 23(1):99–112.
- Cao, Y., L. Hinz, B. Metzke, J. Stein, and A. Holtrop. 2015a. Modeling and mapping fish abundance across Wadeable streams of Illinois, USA, based on landscape-level environmental variables. *Canadian Journal of Fisheries and Aquatic Sciences*:1–16.
- Cao, Y., A. Stodola, S. Douglass, D. Shasteen, and K. Cummings. 2015b. Modelling and mapping the distribution, diversity and abundance of freshwater mussels (Family Unionidae) in Wadeable streams of Illinois, U.S.A. *Freshwater Biology* 60(7):1379–1397.
- Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. *Ecology* 88(11):2783–2792.
- Dubuis, A., J. Pottier, V. Rion, L. Pellissier, J.-P. Theurillat, and A. Guisan. 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions* 17(6):1122–1131.
- Elith, J., and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* 40(1):677–697.
- Ferrier, S., and A. Guisan. 2006. Spatial modelling of biodiversity at the community level. *Journal of applied ecology* 43(3):393–404.



- Ferrier, S., G. Watson, J. Pearce, and M. Drielsma. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity & Conservation* 11(12):2275–2307.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19(4):474–499.
- Guisan, A., and C. Rahbek. 2011. SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography* 38(8):1433–1444.
- Hanski, I., and C. D. Thomas. 1994. Metapopulation dynamics and conservation: A spatially explicit model applied to butterflies. *Biological Conservation* 68(2):167–180.
- He, Y., J. Wang, S. Lek-Ang, and S. Lek. 2010. Predicting assemblages and species richness of endemic fish in the upper Yangtze River. *Science of the Total Environment* 408(19):4211–4220.
- Hinz, L. C. J., and B. Metzke. 2015. Monitoring and Assessment of Aquatic Life in the Kaskaskia River for evaluating IDNR Private Lands Programs: Annual Report 2015. Springfield, IL.
- Hochachka, W. M., R. Caruana, D. Fink, A. Munson, M. Riedewald, D. Sorokina, and S. Kelling. 2007. Data-Mining Discovery of Pattern and Process in Ecological Systems. *The Journal of Wildlife Management* 71(7):2427–2437.
- Huntley, B. 1991. How plants respond to climate change: migration rates, individualism and the consequences for plant communities. *Annals of Botany* 67(suppl):15–22.
- IL-DNR. 2010. Fisheries manual of operation: field sampling protocols for rivers and streams. IL-DNR, Division of Fisheries/Illinois Natural History Survey, 1816 S Oak Street, Champaign, IL 61821.
- Iriondo, J. M., M. J. Albert, and A. Escudero. 2003. Structural equation modelling: an alternative for assessing causal relationships in threatened plant populations. *Biological conservation* 113(3):367–377.
- Iverson, L. R. 1988. Land-use changes in Illinois, USA: The influence of landscape attributes on current and historic land use. *Landscape Ecology* 2(1):45–61.
- Knudby, A., A. Brenning, and E. LeDrew. 2010. New approaches to modelling fish–habitat relationships. *Ecological Modelling* 221(3):503–511.
- Labay, B. J., D. A. Hendrickson, A. Cohen, and K. O. Winemiller. 2015. Can Species Distribution Models Aid Bioassessment when Reference Sites are Lacking? Tests Based on Freshwater Fishes. *Environmental Management* (JUNE).
- Lauenroth, W. K., D. L. Urban, D. P. Coffin, W. J. Parton, H. H. Shugart, T. B. Kirchner, and T. M. Smith. 1993. Modeling vegetation structure–ecosystem process interactions across sites and ecosystems. *Ecological Modelling* 67(1):49–80.

- Lawler, J. J., S. L. Shafer, D. White, P. Kareiva, E. P. Maurer, A. R. Blaustein, and P. J. Bartlein. 2009. Projected climate-induced faunal change in the Western Hemisphere. *Ecology* 90(3):588–597.
- Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. *R News* 2(3):18–22.
- Mateo, R. G., Á. M. Felicísimo, J. Pottier, A. Guisan, and J. Muñoz. 2012. Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS One* 7(3):e32586.
- Mattingly, R. L., E. E. Herricks, and D. M. Johnston. 1993. Channelization and levee construction in Illinois: Review and implications for management. *Environmental Management* 17(6):781–795.
- McIsaac, G. F., and X. Hu. 2004. Net N input and riverine N export from Illinois agricultural watersheds with and without extensive tile drainage. *Biogeochemistry* 70(2):253–273.
- McKenna Jr, J. E., and J. H. Johnson. 2011. Landscape models of Brook Trout abundance and distribution in lotic habitat with field validation. *North American Journal of Fisheries Management* 31(4):742–756.
- Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between – and reconciliation of – “predictive” and “explanatory” models. *Biodiversity & Conservation* 9(5):655–671.
- Old, S. M. 1969. Microclimate, Fire, and Plant Production in an Illinois Prairie. *Ecological Monographs* 39(4):355–384.
- Parker, J., J. Epifanio, A. Casper, and Y. Cao. 2015. The Effects of Improved Water Quality on Fish Assemblages in a Heavily Modified Large River System. *River Research and Applications*.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190(3-4):231–259.
- Piñeiro, G., S. Perelman, J. P. Guerschman, and J. M. Paruelo. 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling* 216(3–4):316–322.
- R Core, T. 2013. R: A language and environment for statistical computing. R Core Team, Vienna, Austria.
- Rehfeldt, G. E., N. L. Crookston, M. V Warwell, and J. S. Evans. 2006. Empirical analyses of plant–climate relationships for the western United States. *International Journal of Plant Sciences* 167(6):1123–1150.
- Sampson, H. 1921. An Ecological Survey of the Prairie Vegetation of Illinois. *Illinois Natural History Survey Bulletin* 13(16).
- Schnier, S., X. Cai, and Y. Cao. 2016. Importance of Natural and Anthropogenic Environmental Factors to Fish Communities of the Fox River in Illinois. *Environmental Management* 57(2):389–411.

- Smith, E. P., and K. A. Rose. 1995. Model goodness-of-fit analysis using regression and related techniques. *Ecological Modelling* 77(1):49–64.
- Steen, P. J., T. G. Zorn, P. W. Seelbach, and J. S. Schaeffer. 2008. Classification tree models for predicting distributions of Michigan stream fish from landscape variables. *Transactions of the American Fisheries Society* 137(4):976–996.
- Williams, J. W., and S. T. Jackson. 2007. Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment* 5(9):475–482.

## **Chapter 3**

### **The Conservation Reserve Program Effects on Stream Fish Diversity in the Kaskaskia River Basin, IL**

#### **Abstract**

Freshwaters support over 40% of fish species diversity, as well as one-third of all vertebrate species. Despite this importance, freshwaters remain one of the most threatened habitats globally, largely due to human impacts. Best management practices (BMPs) have been developed to help mitigate the anthropogenic impacts to which freshwaters are subject. The Conservation Reserve Program (CRP) implements BMPs on agricultural lands with monetary incentives given to landowners willing to participate. Here I use a novel standardization method to assess whether CRP lands affect stream fish species richness. Using a house-neighborhood framework developed from the regional species pool concept, “neighborhoods” were built around a group of fish sampling sites in the Kaskaskia River basin, Illinois. The species richness of the neighborhood was then used to standardize the local species richness. This standardized species richness is then used in a logit regression to assess the effects that CRP and other local habitat variables have on species richness. Proportion of CRP lands at sites ranged from 0% to 45.13%. Using the dredge function within the MuMIn package in R, all possible models were explored.  $R^2$  values were low across all models. None of the best models ( $\Delta AIC < 2$ ) used the proportion of CRP lands as a predictor, instead using various combinations of in-stream habitat features with large substrate consistently being ranked as one of the most important variables affecting species richness. I conclude that my study was unable to detect any major influence from CRP on stream fish species richness.

## **Introduction**

Surface freshwaters are refuge to over 40% of fish species diversity, while only amounting to about 0.01% of water, globally (Lundberg et al. 2000). Rivers, streams, and lakes are not only lifelines for freshwater fish, however. One-third of all vertebrate species are confined to the 0.8% of Earth's surface that is freshwater (Dudgeon et al. 2006). Human activities are having a profound impact on these systems, causing biodiversity loss through a myriad of mechanisms. We continue to search for reasonable and effective management strategies to reverse human caused changes in these important freshwater systems.

Agricultural production is one major anthropogenic source of degradation that has greatly impacted stream ecosystems all over the world. Allan's (2004) review of land use impacts on stream ecosystems noted that as a result of enormous shifts of land use toward agriculture by humans, aquatic ecosystems have suffered. Nonpoint sources of pollution and modifications of natural systems have contributed to decreases in water quality and in habitat availability and complexity (Roth et al. 1996, Richards et al. 1996), and a shift toward a higher proportion of tolerant and non-native species has been seen in biological communities (Lenat and Crawford 1994, Genito et al. 2002). Hydrologic modifications meant to increase agricultural production ultimately resulted in decreased soil quality and loss of farmland due to erosion (Blann et al. 2009), which spurred the creation of several federal conservation programs.

The USDA Conservation Reserve Program (CRP) was implemented as part of the 1985 US Farm Bill to halt the loss of farmland to erosion and to reduce the effects of nutrient input and sedimentation from agriculture on freshwater systems. Best management practices (BMPs) are employed on private lands to achieve these goals, with monetary incentives given to landowners willing to participate. Knight and Boyer (2007) document that the implementation of

conservation programs, such as CRP, can have a detectable effect on water quality and soil condition nationally. Brady (2007) noted that when planned correctly, conservation programs such as CRP show detectable increases in wildlife richness and abundance as well. However, the evidence for such biological changes in aquatic ecosystems becomes much more unclear (Knight and Boyer 2007).

The State of Illinois entered into an agreement with the federal government in 1998 to begin the state-run CRP extension, the Conservation Reserve Enhancement Program (CREP) (FSA 2011). CREP allows the state to offer additional monetary incentives for particular best management practices and longer easements that may be more beneficial to long-term state-specific goals (Bruce 2004). The Illinois CREP has goals of reduced sedimentation, reduced nutrients, 15% increases in bird populations, and 10% increases in native fish and mussel stocks (Illinois Conservation Reserve Enhancement Program 1999).

In 1998, the Illinois River watershed was entered into the original Illinois CREP agreement. The program was expanded in 2010 to include the Kaskaskia River watershed (Illinois Conservation Reserve Enhancement Program 2010), the second largest watershed contained completely within Illinois. While the goals of the program have not changed, very little research has been conducted assessing the effectiveness of CRP or CREP for achieving some of the biological goals defined in the Illinois CREP agreement. Reviewing the annual Illinois CREP reports has produced some documentation of fish-oriented research (Illinois Conservation Reserve Enhancement Program 2004, Carney 2009), though very little. After nearly twenty years of Illinois CREP implementation, a look at whether or not these programs are actually on their way toward reaching the state's fisheries goal stated in the CREP agreement is needed.

It is a complex question to address, though, as many factors at landscape and local scales interact simultaneously to affect species diversity in a particular location. We could simply plot local species richness (SR) with the proportion of CRP lands in the local watershed (PropCRP) as in the equation below.

$$SR = f(PropCRP)$$

This would oversimplify the biology in question, however. We would be left with an unreliable assessment that did not take into account variables impacting richness from different spatial scales due to under-parameterization (Warren and Seifert 2011). At the other end of the spectrum, it is undesirable to incorporate all environmental variables from all spatial scales into the model, as the result would be messy and difficult-to-interpret due to noisy data, many collinear variables, and overfitting. There tends to be an optimum level of complexity at which models operate (Warren and Seifert 2011). Scientists and managers are tasked with adapting methodologies to account for the associated environmental noise that occurs when trying to look at relationships between a number of land use variables and stream condition (Allan 2004). This can be especially difficult when the response variable is biotic, as in my question.

The regional species pool hypothesis (Taylor et al. 1990) states that the biological community in a particular area is molded based upon a number of environmental factors that, over geologic time, cause variations in opportunities for evolution. The assumption is that environmental factors such as climate, historic landscape composition, and soil types have dictated evolution and created a pool of species that best takes advantage of niches within a system. Habitat characteristics at local sites act as a filter to the regional species pool, thus assembling a local community (Poff 1997).

The house-neighborhood framework was developed in line with the regional species pool concept (Merovich et al. 2013), and in this study it is adopted to assess the effect of CRP lands on fish species richness at study sites. A house is defined as a single confluence-to-confluence section of stream and a neighborhood is defined as all waterways connected to the stream within a specified waterway distance. The species found in the neighborhood represent the species pool, as molded by the landscape and other non-local determinants. The species richness of a neighborhood can be defined uniquely for every sampling location. My premise, here, is that that a stream reach within a high diversity neighborhood would likely have a species richness that is greater than a reach in a low diversity neighborhood, regardless of CRP. If the neighborhood richness is used to standardize a sampled species richness, then the result is a richness ratio that is much less affected by the species pool and the non-local landscape factors. By removing the effect of the species pool and looking at the proportion of the species pool that a reach contains, the relationship between how species richness changes due to local factors may be observed more directly. I consider the proportion of CRP in the local watershed to be a local variable, and am using CRP as a surrogate for CREP. The justification for this surrogacy is that CREP lands must be CRP lands prior to CREP enrollment, and CREP uses CRP practices. Using the neighborhood-standardized species richness, developed from individual species distribution models (SDMs), I aimed to investigate the effect of CRP lands and local habitat on fish species richness in wadeable tributaries of the Kaskaskia River.



## Methods

### *Field Sampling*

The Stream Ecology Research Group of the Illinois Natural History Survey (<http://www.inhs.illinois.edu/programs/aquatic/stream-ecosystems/stream-ecosystems/>) conducts fisheries sampling in wadeable tributaries of the Kaskaskia River using electric seine and backpack electrofishing (Hinz and Metzke 2015). Site resolution for the dataset is confluence-to-confluence sections of stream. Thirty to thirty-five sites were sampled during base-flow conditions in the summer months of 2013, 2014, and 2015 using the Illinois Department of Natural Resources standard electrofishing protocols (IL-DNR 2010). Sampling reaches at each site were at least 100 meters, and ideally 20x the stream width. After each sample was completed, fish were identified to species, recorded, and returned to the stream. A portion of these sites were used in building SDMs, and these sites were not included as study sites for this analysis. Forty-seven remaining sites made up my dataset for this study.

In addition to fish samples, other physical and chemical attributes were measured and recorded (Hinz and Metzke 2015). Qualitative habitat assessments were done at each site using the Illinois Habitat Index (Sass et al. 2007) and Qualitative Habitat Evaluation Index (Rankin 2006). Habitat was assessed by the same crew member at every site to avoid subjective discrepancies in qualitative habitat description. In situ nutrient tests for reactive phosphorus, nitrates, ammonia, and turbidity were conducted using a HACH DR 900 Multiparameter Handheld Colorimeter (<http://www.hach.com>). Dissolved oxygen, pH, and conductivity were recorded from moving water in the middle of the channel using a HACH HQ40 Multimeter. If moving water was not available, probes were either swirled to create moving water or moved to another portion of the stream that had faster flowing water.

### *Building Neighborhoods*

My objective was to use previously built, reliable SDMs to build neighborhood richness around the 46 sites used for this study. This would then allow me to standardize measured richness of the 46 sites and use regression to examine the relationship between CRP and standardized richness. SDMs for 52 species were built using random forests classification (Breiman 2001) within R (R Core Team 2013), using the package “randomForest” (Liaw and Wiener 2002).

Network Analyst in ArcMap (ESRI 2013) was used to build an origin-destination cost matrix between all confluence-to-confluence reaches in the Kaskaskia River basin. I identified all stream arcs within stream distances of 5km, 10km, 15km, and 20km from each sampling location in the dataset. Each distance from the sampling location indicated a neighborhood. Current literature suggests that in aquatic systems the regional species pool for both invertebrates (Sundermann et al. 2011) and fish (Stoll et al. 2013) is most constraining at smaller distances (~5km). Neighborhood species richness was calculated, using the presence/absence predictions from my SDMs, for every neighborhood distance around each of the 46 sampling locations.

### *Regression and Dredging*

Species richness from each of a site’s neighborhoods was used to standardize the local, measured richness and remove the effect of the species pool and limit landscape derived environmental noise from the dataset. This resulted in four separate standardized species richness values, one each for 5km, 10km, 15km, and 20km neighborhood distances. Linear modeling was used to explore the relationship between CRP in the local watershed and these standardized species richness values. The standardized values were transformed using the logit transformation

(Warton and Hui 2011) to avoid issues concerning the use of proportional variables as dependent variables in regression. The logit transformation is expressed as

$$\text{logit}(x) = \log \left( \frac{x}{1-x} \right)$$

Using the dredge function of the MuMIn package (Barton 2013) in R, I explored all possible models explaining the dependence of the standardized, transformed richness value on the proportion of CRP lands in the immediate watershed as well as a select group of local habitat variables. I also examined variable importance taken from dredge function, where the importance for a particular variable is averaged over the models in which that variable was included. The equation below shows my complete model:

$$\text{logit} \left( \frac{SR}{NSR} \right) = f ( \text{PropCRP} + \text{LargeSubstr} + \text{PropRiffle} + \text{Shading} + \text{LWD} + \text{Volume} + \text{IHI} )$$

where SR is the local, measured species richness; NSR is the predicted neighborhood species richness; and PropCRP is the proportion of CRP lands in the immediate watershed. The rest of the variables come from the qualitative habitat assessment, where LargeSubstr is a qualitative measure of the proportion of large substrates in the sampled reach; PropRiffle is the proportion of stream segments classified as riffles in the sampled reach; Shading is the proportion of stream that is shaded by riparian vegetation and trees; LWD is a qualitative measure of the amount of large and aggregate woody debris in the sampled reach; Volume is the average depth of the sampled reach x average width of the reach x the length of the reach; and IHI is the score given to the sampling reach via the Illinois Habitat Index (Sass et al. 2007).

## Results

The 46 sites used in this study were well distributed throughout the Kaskaskia River basin, from tributaries of the lower reaches near the Mississippi River to headwaters in east-central Illinois (Figure 3.1). Proportion of CRP lands in the local watershed of these sites ranged from 0% to 45.13%, and a majority of sites existed in the lower portion of that range (Figure 3.2). This reflected the overall pattern of CRP enrollment in the basin in that higher proportions of CRP were not common. When HUC-8 watersheds within the Kaskaskia River basin are considered, the Middle Kaskaskia HUC-8 shows a higher average proportion of CRP lands, while the Shoal Creek HUC-8 shows the highest average species richness (Table 3.1). The Shoal Creek watershed is considered a higher quality watershed in this basin (i.e. higher diversity), and this points out the need for a consideration of species pool connectivity in my assessment. The Shoal Creek HUC-8 was also only represented by 5 samples in this study (Figure 3.3).

Standardized species richness acted as one would expect, with the proportion of species from the neighborhood represented at a site decreasing as neighborhood distance increased (Table 3.2). When observed richness was plotted against the proportion of CRP land in local watersheds, a negative slope and poor  $R^2$  resulted [ $SR = -2.4416(PropCRP) + 11.649$ ,  $R^2 = 0.0029$ ,  $p = 0.691$ ] (Figure 3.4). Plotting the standardized richness for each neighborhood distance with the proportion of CRP lands in local watersheds changed very little (Figure 3.5), suggesting there are many more factors affecting species richness than CRP lands alone. After incorporating local habitat variables, the best models ( $\Delta AIC < 2$ ) at each neighborhood distance did not take the proportion of CRP lands as a predictor (Table 3.3) which reinforces this point. All of the best models use some combination of in-stream habitat features.  $R^2$  values were low

across all models. This is expected, however, given the wide seasonal and inter-annual variations exhibited in natural systems and the multiple-year data collection that took place for this study.

Variable importance, averaged over all models that a particular variable was included in, shows that large substrate was the most important variable in three of the four neighborhood distances, with water volume becoming the most important variable at the 15km distance (Table 3.4). Over all neighborhood distances, CRP lands were consistently at the low end of the importance table, while local, in-stream habitat features were always much more important for determining richness.

## **Discussion**

My study set out to investigate whether CRP lands are influencing stream fish diversity. I was unable to detect a major influence of CRP lands on stream fish diversity, and instead my results suggest that in-stream habitat is a much more important factor influencing community composition at this time. It is possible, however, that the effects of CRP lands on stream fish diversity are so small or hidden that I was unable to uncover it using my methodology. My results are contrasted by studies involving CRP lands and birds, where mostly positive biological responses to BMPs were seen (Johnson and Schwartz 1993, Johnson and Igl 1995, King and Savidge 1995, Best et al. 1997, Coppedge et al. 2004). My results are similar to studies from the water quality realm, though, where mixed results have occurred (Dosskey 2001, Brueggen-Boman et al. 2015) and BMPs have struggled to show positive changes consistently.

Water quality is certainly a concern in fish management (Meador and Goldstein 2003). My results suggest that CRP alone is not enough to increase stream fish species richness. This

falls within the narrative described by other studies (Karr and Dudley 1981, Feld et al. 2011), where it is emphasized that there are many aspects of ecosystems to consider for successful community restoration. While BMPs that are designed to help water quality, such as those in CRP, are of benefit to aquatic systems (Brady 2007), the degree to which they can be effective may rely on their scale of implementation and on their interconnectedness with other beneficial practices such as in-stream restoration. Brueggen-Bowman et al (2015) suggested that BMPs might offer better results if implemented strategically in target areas. The importance here is that CRP is not a program that is strategically designed to be cooperative with other management activities affecting fish. It does have a competitive enrollment system based on an environmental benefits index, which makes some attempt toward giving more value to lands in predefined water quality zones or wildlife priority zones (FSA 2013). This is not a criterion that is necessary to meet for the program, though. Properties outside these zones are still eligible for CRP contracts, and lands inside these zones can still fail to meet other CRP criteria which could result in exclusion from the program.

Tile drainage systems are employed extensively throughout Illinois and the Kaskaskia basin. McIsaac and Hu (2004) found linkage between tile drained lands and a greater contribution of nutrient input into the Mississippi River system. Tile drainage basically allows for the bypass of many on-land management activities via fast, subsurface flow. With the scale of agriculture in the Midwest (Benke and Cushing 2005, NASS 2012), and particularly in areas such as the Kaskaskia basin where tile drainage systems are employed heavily, it is possible that CRP is not implemented on a scale large enough or focused enough to break necessary thresholds and impact fish diversity in any detectable amount. CRP totals over 860,000 acres in Illinois (FSA 2015), but that accounts for less than 2.5% of the land in Illinois while farmland

occupies almost 27 million acres in the state (NASS 2012) or roughly 73%. As currently applied, any small effect that CRP may be able to have is likely to be diluted by the effect of an enormous amount of anthropogenic disturbance. Both the inability of my neighborhood standardized richness to capture any noticeable change compared to observed species richness and the heavily weighted importance of in-stream habitat variables support this idea. It suggests that water quality conditions are either not what is limiting stream fish species richness in the Kaskaskia River basin or that the water quality effects from CRP and other voluntary BMP programs are simply not great enough to affect richness. A more rigorously targeted approach for CRP to better cooperate with other voluntary BMP programs may be beneficial to address this concern, but future research and monitoring are needed.

Other matters within the Conservation Reserve Program itself may have contributed to my inability to detect much of an effect from CRP lands on fish diversity. My methods lump all CRP practices in as one variable, but there are 44 practices in total within the program that vary widely in application and effect. A similar study to mine, where CRP is broken down by practice type and where local habitat variability is controlled for, could be useful in determining whether CRP lands are having effects on fisheries at all. A study such as this would also allow for the examination of the Illinois CREP program and its eligible practices to determine whether those practices being included were, in fact, the practices capable of achieving fisheries goals.

Additionally, knowing the length of time a parcel was enrolled in CRP would be very valuable, as it is documented that BMP vegetation age structure can impact their effect (Broadmeadow and Nisbet 2004). This information could be included in the model to assess whether there is a lag time in the response to CRP from fish communities. Unfortunately I was unable to obtain data regarding length of time a particular piece of land was enrolled in CRP, which may have been a

source of unexplained variability. Records were kept for the most current CRP contract for a parcel, but records on whether that land was enrolled in CRP years prior were unavailable. In other words, for lands with a new contract last year, there was no way of knowing whether that land was enrolled in CRP for the 15 years prior or not.

Again, my study was unable to detect a major effect of CRP on stream fish species richness. I believe the most likely scenario at play here is that water and habitat quality has been so degraded by agricultural activity in this region that CRP is not implemented in a way or on a scale that can affect species richness noticeably. Additionally, the lack of a practice-by-practice based focus in my study likely contributed to a decreased ability to recognize impacts. Future research and the additional data mentioned previously could help to decipher this to a better degree.



## Tables and Figures

**Table 3.1.** Average proportion CRP and average species richness for the four HUC-8 watersheds within the Kaskaskia River basin

<i>HUC-8</i>	<i>Upper Kaskaskia</i>	<i>Middle Kaskaskia</i>	<i>Shoal Creek</i>	<i>Lower Kaskaskia</i>
Proportion CRP	0.068987	0.147289	0.028348	0.080424
Species Richness	11.16667	10.58333	15	11.18182

**Table 3.2.** Summary table of observed and standardized richness ranges.

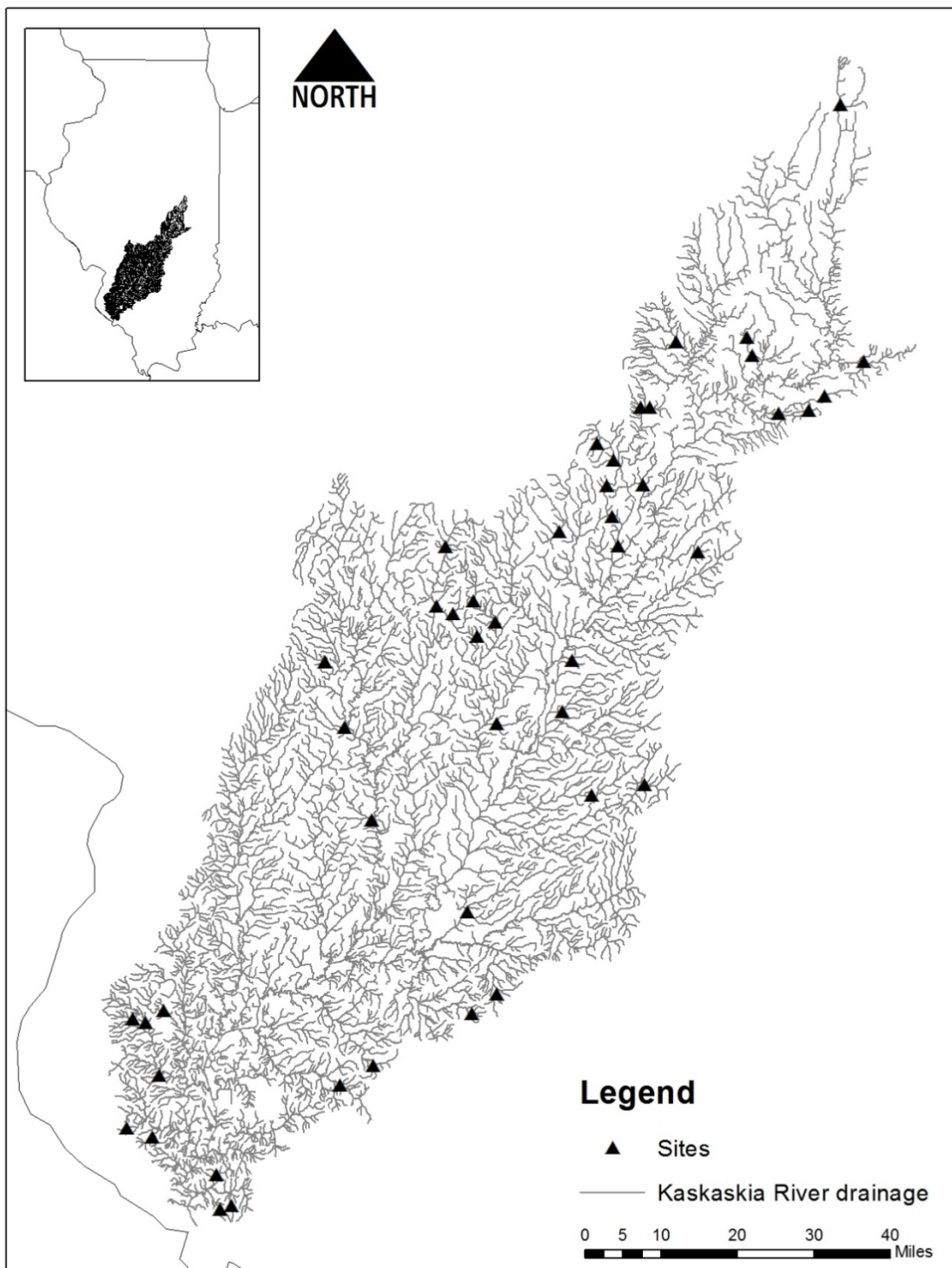
<i>Neighborhood Standardized Richness</i>	<i>Minimum</i>	<i>Mean</i>	<i>Maximum</i>
Observed Richness	1	11.435	20
5km Standardized	0.077	0.637	1.000
10km Standardized	0.062	0.503	0.937
15km Standardized	0.053	0.436	0.769
20km Standardized	0.053	0.392	0.739

**Table 3.3.** Top model ( $\Delta AIC < 2$ ) produced at each neighborhood distance using the dredge function of the MuMIn package in R.

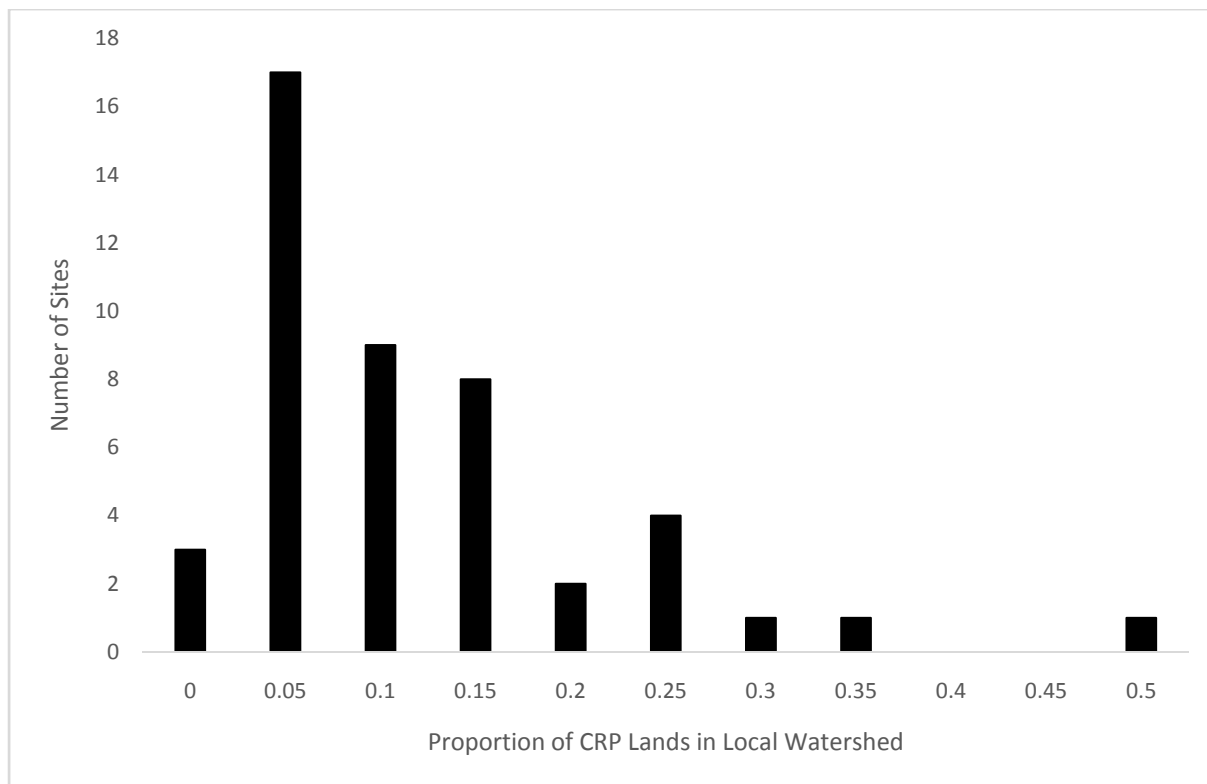
<i>Model</i>	<i>Intercept</i>	<i>IHI</i>	<i>Large Substrate</i>	<i>Prop Riffle</i>	<i>Prop CRP</i>	<i>Shading</i>	<i>LWD</i>	<i>Volume</i>	<i>R<sup>2</sup></i>
Logit(SR/5km)	2.21867	NA	0.064590	NA	NA	NA	-0.2223	NA	0.2367
	1.88783	NA	-0.085142	NA	NA	NA	NA	NA	0.1799
Logit(SR/10km)	-0.8462	NA	0.126454	NA	NA	NA	NA	0.00149	0.2195
	-0.4863	NA	0.118702	NA	NA	NA	NA	NA	0.1531
	-0.3029	NA	NA	NA	NA	NA	NA	0.00153	0.1487
Logit(SR/15km)	-0.9932	NA	0.091791	NA	NA	NA	NA	0.00129	0.2050
	-0.5927	NA	NA	NA	NA	NA	NA	0.00131	0.1522
	-0.6802	NA	0.085050	NA	NA	NA	NA	NA	0.1341
Logit(SR/20km)	-0.8650	NA	0.079543	NA	NA	NA	NA	NA	0.1247
	-1.0812	NA	0.084200	NA	NA	NA	NA	0.00089	0.1655
	-0.7091	NA	NA	NA	NA	NA	NA	0.00090	0.1164
	-0.9553	0.0246	NA	NA	NA	NA	NA	NA	0.0915

**Table 3.4.** Variable importance of each variable included in the dredge function. Values represent the average importance of the variable over all of the models in which that variable was included. Darker greyscale indicates a more important variable in that model.

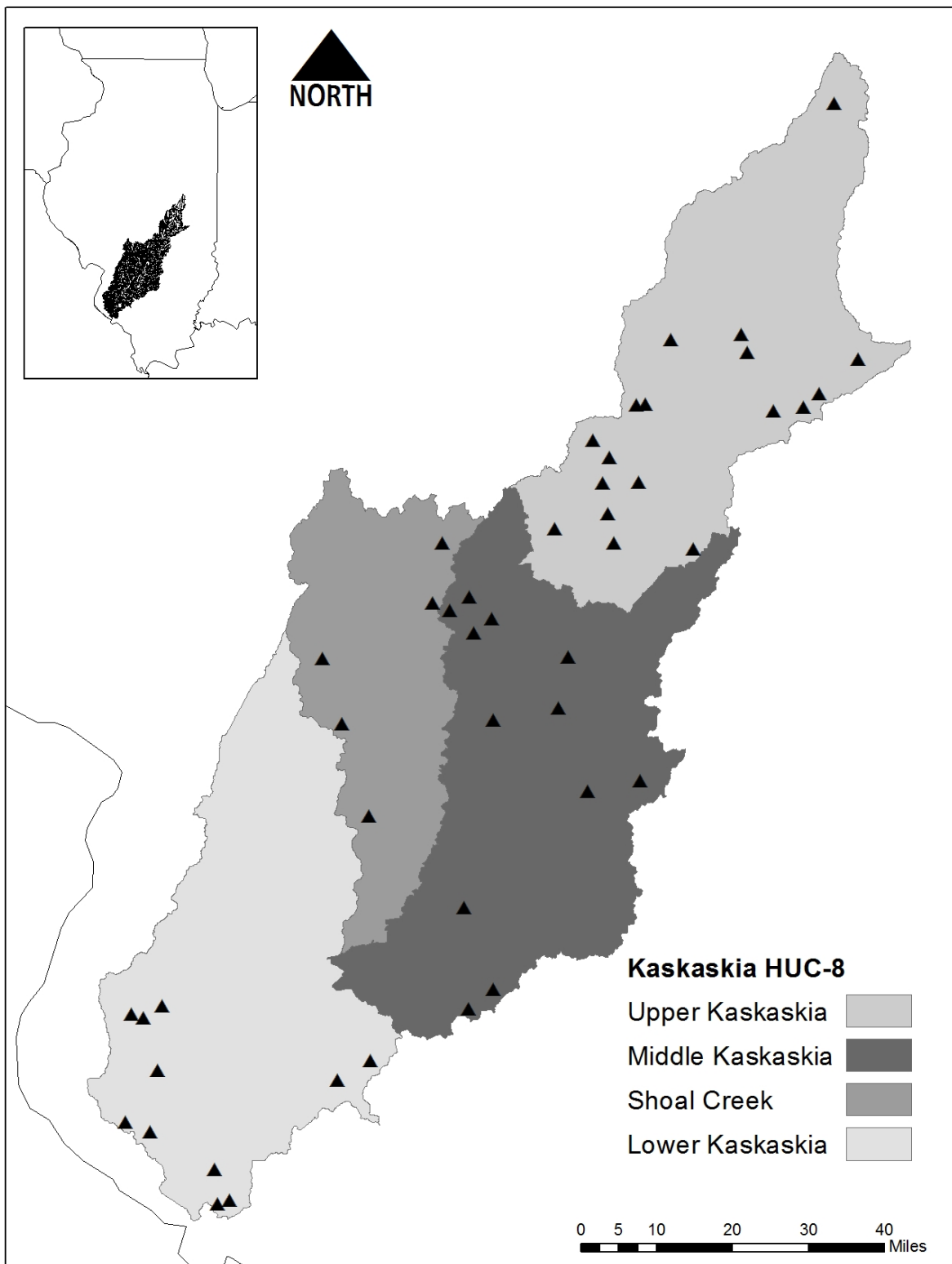
<i>Model</i>	<i>Large Substrate</i>	<i>Volume</i>	<i>IHI</i>	<i>Shading</i>	<i>Woody Debris</i>	<i>Proportion CRP</i>	<i>Proportion Riffles</i>
Logit(SR/5km)	0.781	0.231	0.250	0.241	0.587	0.227	0.000
Logit(SR/10km)	0.686	0.685	0.238	0.238	0.241	0.230	0.000
Logit(SR/15km)	0.546	0.695	0.268	0.239	0.238	0.224	0.000
Logit(SR/20km)	0.526	0.495	0.311	0.253	0.253	0.239	0.000



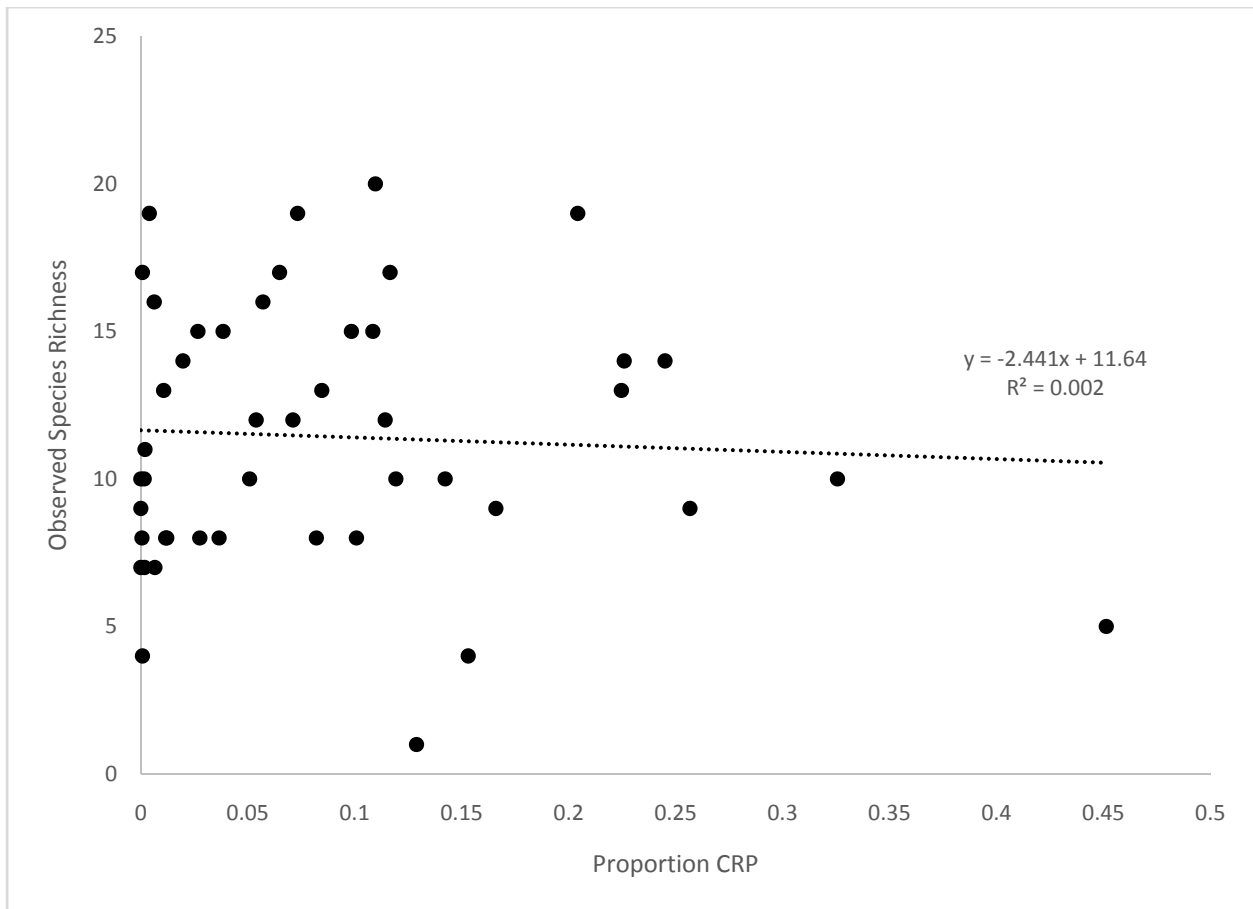
**Figure 3.1.** Spatial distribution of the 46 sites used in this study.



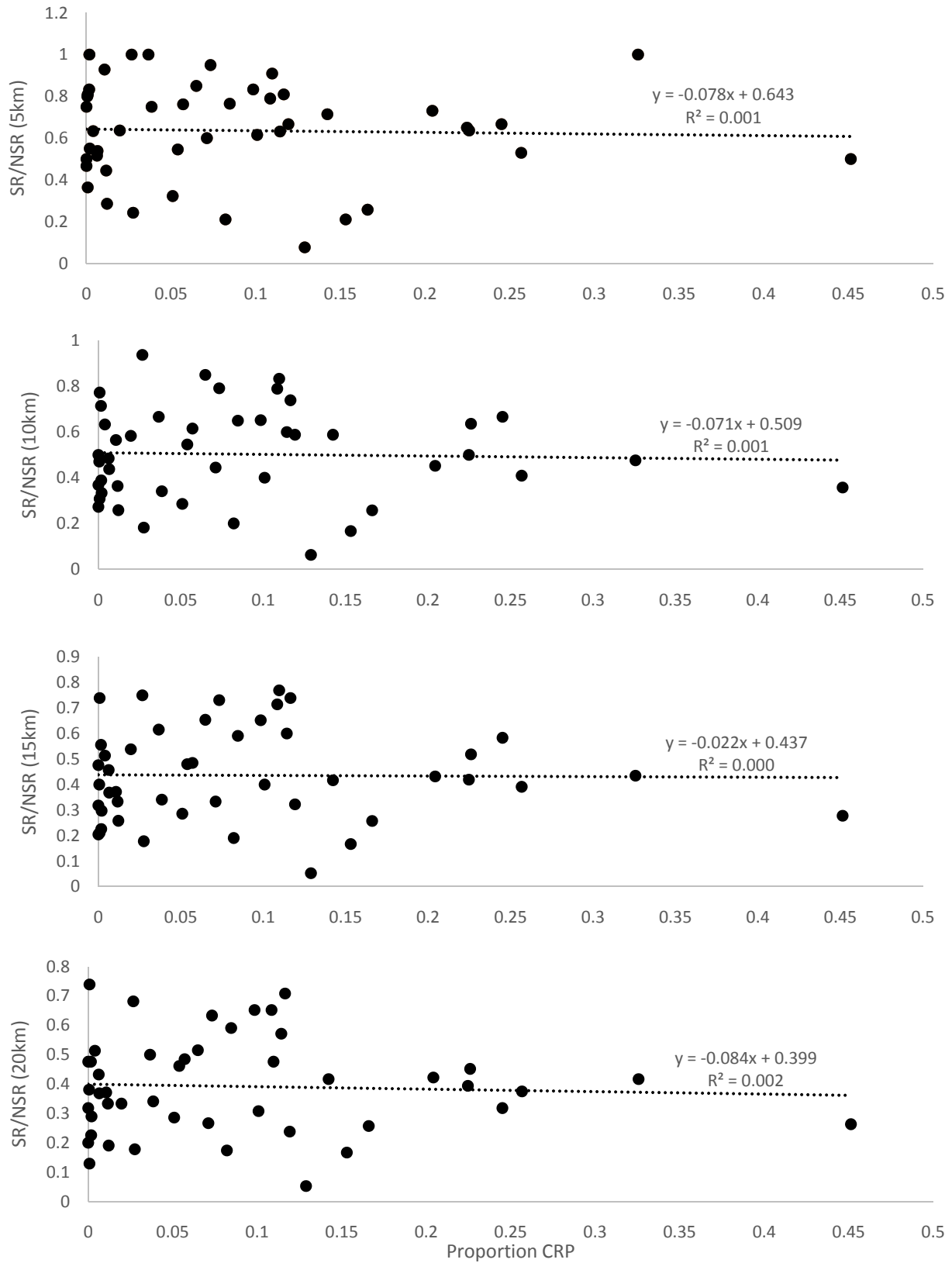
**Figure 3.2.** Frequency histogram reflecting distribution of CRP across sites in this study. This trend is seen over most of the Kaskaskia River Basin.



**Figure 3.3.** Spatial configuration of Kaskaskia River Basin HUC-8 watersheds.



**Figure 3.4.** Observed richness plotted versus proportion CRP.



**Figure 3.5.** Standardized richness plotted versus proportion CRP for every neighborhood distance.

## References

- Allan, J. D. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35(1):257–284.
- Barton, K. 2013. MuMIn: Multi-model inference.
- Benke, A. C., and C. E. Cushing. 2005. *Rivers of North America*. Academic Press, Burlington, MA.
- Best, L. B., H. Campa, K. E. Kemp, R. J. Robel, M. R. Ryan, J. A. Savidge, H. P. Weeks, and S. R. Winterstein. 1997. Bird Abundance and Nesting in CRP Fields and Cropland in the Midwest: A Regional Approach. *Wildlife Society Bulletin (1973-2006)* 25(4):864–877.
- Blann, K. L., J. L. Anderson, G. R. Sands, and B. Vondracek. 2009. Effects of Agricultural Drainage on Aquatic Ecosystems: A Review. *Critical Reviews in Environmental Science and Technology* 39(11):909–1001.
- Brady, S. J. 2007. Effects of Cropland Conservation Practices on Fish and Wildlife Habitat.
- Breiman, L. 2001. Random Forests. *Journal of Machine Learning* 45(1):5–32.
- Broadmeadow, S., and T. R. Nisbet. 2004. The effects of riparian forest management on the freshwater environment: a literature review of best management practice. *Hydrology and Earth System Sciences* 8(3):286–305.
- Bruce, D. 2004. Illinois Conservation Reserve Enhancement Program (CREP): A Model for Watershed Restoration. Pages 95–99 in L. E. Riley, R. K. Dumroese, and T. D. Landis, editors. *National Proceedings: Forest and Conservation Nursery Associations-2003*. US Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Brueggen-boman, T. R., S. Choi, and J. L. Bouldin. 2015. Response of Water-Quality Indicators to the Implementation of Best-Management Practices in the Upper Strawberry River Watershed , Arkansas:697–713.
- Carney, D. 2009. Smallmouth Bass Assessment in Sugar Creek, Bellrose Nature Preserve, Logan County, IL. Illinois Department of Natural Resources Memorandum. Springfield, IL.
- Coppedge, B. R., D. M. Engle, R. E. Masters, and M. S. Gregory. 2004. Predicting juniper encroachment and CRP effects on avian community dynamics in southern mixed-grass prairie, USA. *Biological Conservation* 115:431–441.
- Dosskey, M. G. 2001. Toward quantifying water pollution abatement in response to installing buffers on crop land. *Environmental Management* 28(5):577–598.
- Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. L&eacute;vy&ccirc;que, R. J. Naiman, A.-H. Prieur-Richard, D. Soto, M. L. J. Stiassny, and C. A. Sullivan. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* 81(02):163–182.
- ESRI. 2013. ArcGIS Desktop. Environmental Systems Reserach Institute, Redlands, CA.



- Feld, C. K., S. Birk, D. C. Bradley, D. Hering, J. Kail, A. Marzin, A. Melcher, D. Nemitz, M. L. Pedersen, F. Pletterbauer, D. Pont, P. F. M. Verdonshot, and N. Friberg. 2011. *From Natural to Degraded Rivers and Back Again*. Advances in Ecological Research, 1st edition. Elsevier Ltd.
- FSA. 2011. Fact Sheet.
- FSA. 2013. EBI Fact Sheet.
- FSA. 2015. CRP Monthly Summary December 2015.
- Genito, D., W. J. Gburek, and A. N. Sharpley. 2002. Response of stream macroinvertebrates to agricultural land cover in a small watershed. *Journal of Freshwater Ecology* 17(1):109–119.
- Hinz, L. C. J., and B. Metzke. 2015. *Monitoring and Assessment of Aquatic Life in the Kaskaskia River for evaluating IDNR Private Lands Programs: Annual Report 2015*. Springfield, IL.
- IL-DNR. 2010. Fisheries manual of operation: field sampling protocols for rivers and streams. IL-DNR, Division of Fisheries/Illinois Natural History Survey, 1816 S Oak Street, Champaign, IL 61821.
- Illinois Conservation Reserve Enhancement Program. 1999. *Assessment of the Illinois River Conservation Reserve Enhancement Program*. Springfield, IL.
- Illinois Conservation Reserve Enhancement Program. 2004. *Annual Report for reporting period October 2002 through September 2003*. Springfield, IL.
- Illinois Conservation Reserve Enhancement Program. 2010. *Annual Report for reporting period October 2009 through September 2010*. Springfield, IL.
- Johnson, D. H., and L. D. Igl. 1995. Contributions of the Conservation Reserve Program to Populations of Breeding Birds in North Dakota. *The Wilson Bulletin* 107(4):709–718.
- Johnson, D. H., and M. D. Schwartz. 1993. The Conservation Reserve Program and Grassland Birds. *Conservation Biology* 7(4):934–937.
- Karr, J. R., and D. R. Dudley. 1981. Ecological perspective on water quality goals. *Environmental management* 5(1):55–68.
- King, J. W., and J. A. Savidge. 1995. Effects of the Conservation Reserve Program on Wildlife in Southeast Nebraska. *Wildlife Society Bulletin (1973-2006)* 23(3):377–385.
- Knight, S., and K. Boyer. 2007. *Effects of Conservation Practices on Aquatic Habitats and Fauna*.
- Lenat, D. R., and J. K. Crawford. 1994. Effects of land use on water quality and aquatic biota of three North Carolina Piedmont streams. *Hydrobiologia* 294(3):185–199.
- Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. *R News* 2(3):18–22.

- Lundberg, J. G., M. Kottelat, G. R. Smith, M. L. J. Stiassny, and A. C. Gill. 2000. So Many Fishes, So Little Time: An Overview of Recent Ichthyological Discovery in Continental Waters. *Annals of the Missouri Botanical Garden* 87(1):26–62.
- McIsaac, G. F., and X. Hu. 2004. Net N input and riverine N export from Illinois agricultural watersheds with and without extensive tile drainage. *Biogeochemistry* 70(2):253–273.
- Meador, M. R., and R. M. Goldstein. 2003. Assessing water quality at large geographic scales: relations among land use, water physicochemistry, riparian condition, and fish community structure. *Environmental Management* 31(4):504–517.
- Merovich, G., T. Petty, M. Strager, and J. B. Fulton. 2013. Hierarchical classification of stream condition : A house-neighborhood framework for establishing conservation priorities in complex riverscapes. *Freshwater Science* 32(3):874–891.
- NASS, U. 2012. Census of agriculture. US Department of Agriculture, National Agricultural Statistics Service, Washington, DC.
- Poff, N. L. 1997. Landscape Filters and Species Traits: Towards Mechanistic Understanding and Prediction in Stream Ecology. *Journal of the North American Benthological Society* 16(2):391–409.
- R Core, T. 2013. R: A language and environment for statistical computing. R Core Team, Vienna, Austria.
- Rankin, E. T. 2006. Methods for Assessing Habitat in Flowing Waters : Using the Qualitative Habitat Evaluation Index ( QHEI ). Ohio EPA Technical Bulletin.
- Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influences on stream habitats and biota. *Canadian Journal of Fisheries and Aquatic Sciences* 53(S1):295–311.
- Roth, N. E., J. D. Allan, and D. L. Erickson. 1996. Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landscape ecology* 11(3):141–156.
- Sass, L., L. Hinz, and J. Epifanio. 2007. Developing a multi-metric habitat index for wadeable streams in Illinois. *Environmental Protection* 2007(June):15.
- Stoll, S., A. Sundermann, A. W. Lorenz, J. Kail, and P. Haase. 2013. Small and impoverished regional species pools constrain colonisation of restored river reaches by fishes. *Freshwater Biology* 58(4):664–674.
- Sundermann, A., S. Stoll, and P. Haase. 2011. River restoration success depends on the species pool of the immediate surroundings. *Ecological Applications* 21(6):1962–1971.
- Taylor, D. R., L. W. Aarssen, and C. Loehle. 1990. On the relationship between r/K selection and environmental carrying capacity: A new habitat templet for plant life history strategies. *Oikos* 58(2):239–250.
- Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21(2):335–342.

Warton, D. I., and F. K. C. Hui. 2011. The arcsine is asinine: the analysis of proportions in ecology. *Ecology* 92(1):3–10.

## **Chapter 4**

### **Summary**

Chapter 1 provided a review of freshwater ecosystems and best management practices. I described how anthropogenic disturbances have made major impacts on stream ecosystems (Allan 2004). These have come in forms encompassing all aspects of agriculture (Blann et al. 2009), urbanization (Paul and Meyer 2001), and deforestation (Ice 2004). Hydrologic and morphological alterations, as well as sediment and nutrient loading, have been the major mechanisms by which humans have degraded water quality and habitat in streams (Feld et al. 2011). While it can sometimes be easy to identify and solve point source pollution, discovering and mitigating non-point source pollution presents a much bigger challenge. Throughout the 20<sup>th</sup> and 21<sup>st</sup> centuries, best management practices (BMPs) were created and refined to address non-point source pollution (Ice 2004). BMPs have been widely implemented, but in terms of reaching their water quality or biological goals their results have been fairly inconsistent (Allan 2004, Tomer and Locke 2011). The increasing discussion of incorporating connectivity into conservation planning has brought about new ideas to help BMPs achieve their goals more consistently (Brueggen-boman et al. 2015, Holmes et al. 2016).

The Conservation Reserve Program (CRP) uses BMPs, on a voluntary basis, in an attempt to decrease anthropogenic impacts on water and soil quality. Some state variations of CRP include biological goals as well. These are often aimed at birds and fish, such as in the Illinois Conservation Reserve Enhancement Program (Illinois Conservation Reserve Enhancement Program 1999). These goals built the foundations for the question in my study, which was whether CRP could affect stream fish species richness. Studies have shown the reliance of fish communities on connectivity to a regional species pool, where species dispersal

limitations from source populations can limit colonization of newly restored sites (Stoll et al. 2013, 2014). I intended to use a novel approach to address this connectivity in my study. I standardized the local species richness by the regional species pool, via a house-neighborhood framework. By doing this, species pool effects and non-local environmental effects were accounted for my study design, which explored the relationship between local habitat variables and species richness. In order to do this, it was necessary to model species distributions in order to obtain the necessary amount of data to provide estimates of species pools.

In Chapter 2, I used Random Forests classification (Breiman 2001) to model presence and absence of 64 species found in the Kaskaskia River basin. Of these, 52 species models exceeded my threshold for model performance of  $TSS > 0.2$ , indicating they performed adequately. No threshold is recommended by the literature, therefore this threshold was set by me in order to balance acceptable model performance with the number of usable species distribution models. Models were then stacked to look at species richness for stream reaches across the entire Kaskaskia River basin. The predicted species richness compared excellently to observed species richness with a Theil's Inequality Coefficient of  $U = 0.011$ . This provided a necessary foundation for Chapter 3 by providing information on species distributions at a finer resolution than sampling alone could have provided in a cost-efficient manner. This fine resolution species distribution information was used to build neighborhood species pools.

In Chapter 3, the predicted species richness was used to build neighborhoods around a group of sites within the basin. The neighborhood richness for each site was then used as a standardization. My study was ultimately unable to decipher any major influence from CRP on species richness, as it was not used as a factor in any of the top models ( $\Delta AIC < 2$ ) and ranked very low, relatively, in variable importance. Local habitat variables such as large substrate,

volume, and woody debris were consistently measured as the most important factors influencing species richness. Chapter 3 highlights the importance of local habitat variables to species richness. In addition, it makes a case for the continued improvement of BMPs and their implementation scheme across the Kaskaskia River basin. It would make sense that CRP would have a noticeable, positive effect on species richness, as the local water quality conditions were likely to improve surrounding the CRP site. However, my inability to capture an effect from CRP on species richness means that water quality conditions may not have sufficiently improved or that other factors at these sites were just overwhelmingly more important.

Additionally, the shortcomings of my study, also described in Chapter 3, could have contributed to the lack of noticeable effect from CRP. Length-of-participation data or a more practice specific study could have potentially improved my ability to find a larger effect. Vegetation age structure has been shown to impact the effectiveness of streamside BMPs (Broadmeadow and Nisbet 2004), making data describing the length of time a BMP was implemented on a particular piece of land valuable. Unfortunately this data was not available. Given the wealth of information documenting the difficulty of finding a positive effect in aquatic ecosystems from BMPs, and that the targeting approach within the program itself has not yet been improved upon, I am not very surprised that my study had difficulty finding a positive effect from CRP. My inability to find an effect, though, highlights that an investigation into the coordination between CRP sites and other voluntary BMP programs may be needed. More rigorous monitoring could help identify areas for improvement, and a better targeted approach to CRP could help to overcome the enormous amount of agricultural impacts in the basin.

Overall, my study used a combination of sampling, GIS, and modeled data to investigate the effects of a voluntary private lands conservation program on species richness in the

Kaskaskia River basin. It is an example of how sampling and modeling can be applied simultaneously to answer a question, and shows the importance of long term datasets and data collaboration to the advancement of our field. Continued collection of long term water quality, fisheries, and BMP length-of-implementation data is imperative to answering questions such as mine. Ongoing monitoring and rigorous evaluation of BMPs and their implementation strategies is likely to be necessary to improve the condition of our streams and rivers in any meaningful way.

### References

- Allan, J. D. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35(1):257–284.
- Blann, K. L., J. L. Anderson, G. R. Sands, and B. Vondracek. 2009. Effects of Agricultural Drainage on Aquatic Ecosystems: A Review. *Critical Reviews in Environmental Science and Technology* 39(11):909–1001.
- Breiman, L. 2001. Random Forests. *Journal of Machine Learning* 45(1):5–32.
- Broadmeadow, S., and T. R. Nisbet. 2004. The effects of riparian forest management on the freshwater environment: a literature review of best management practice. *Hydrology and Earth System Sciences* 8(3):286–305.
- Brueggen-boman, T. R., S. Choi, and J. L. Bouldin. 2015. Response of Water-Quality Indicators to the Implementation of Best-Management Practices in the Upper Strawberry River Watershed , Arkansas:697–713.
- Feld, C. K., S. Birk, D. C. Bradley, D. Hering, J. Kail, A. Marzin, A. Melcher, D. Nemitz, M. L. Pedersen, F. Pletterbauer, D. Pont, P. F. M. Verdonshot, and N. Friberg. 2011. From Natural to Degraded Rivers and Back Again. *Advances in Ecological Research*, 1st edition. Elsevier Ltd.
- Holmes, R., D. G. Armanini, and A. G. Yates. 2016. Effects of Best Management Practice on Ecological Condition : Does Location Matter ? *Environmental Management*.
- Ice, G. 2004. History of Innovative Best Management Practice Development and its Role in Addressing Water Quality Limited Waterbodies. *Journal of Environmental Engineering* 130(6):684–689.
- Illinois Conservation Reserve Enhancement Program. 1999. Assessment of the Illinois River Conservation Reserve Enhancement Program. Springfield, IL.

- Paul, M. J., and J. L. Meyer. 2001. Streams in the Urban Landscape. *Annual Review of Ecology and Systematics* 32(1):333–365.
- Stoll, S., J. Kail, A. W. Lorenz, A. Sundermann, and P. Haase. 2014. The importance of the regional species pool, ecological species traits and local habitat conditions for the colonization of restored river reaches by fish. *PloS one* 9(1):1–10.
- Stoll, S., A. Sundermann, A. W. Lorenz, J. Kail, and P. Haase. 2013. Small and impoverished regional species pools constrain colonisation of restored river reaches by fishes. *Freshwater Biology* 58(4):664–674.
- Tomer, M. D., and M. A. Locke. 2011. The challenge of documenting water quality benefits of conservation practices: a review of USDA-ARSs conservation effects assessment project watershed studies. *Water Science and Technology* 64(1):300.